ULTRA-LOW-POWER, LOW-VOLTAGE DIGITAL CIRCUITS FOR BIOMEDICAL SENSOR NODES

MAHMOOD KHAYATZADEH

(B.SC., AMIRKABIR UNIVERSITY OF TECHNOLOGY, IRAN)

(M.SC., AMIRKABIR UNIVERSITY OF TECHNOLOGY, IRAN)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF

ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2013

To my parents for their unbounded and unconditional love and support

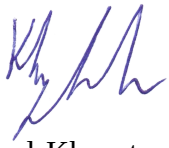
And

To my forever wonderful love, Azadeh

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Mahmood Khayatzadeh

Acknowledgments

I would like to express my endless appreciation to my supervisor, Professor Dr. Yong Lian. I am grateful for his valuable guidance and encouragement throughout my PhD work. He let me examine and experience the ideas that came into my mind and gave me time to explore them. Without his guidance, valuable comments, and continuous support, this work would not be possible.

I would also like to express my appreciation to Dr. Chun-Huat Heng, Dr. Massimo Alioto, Professor Dr. Dennis Sylvester, and Professor Dr. David Blaauw for their inspiring support in my research.

I am thankful to my friends and colleagues in the signal processing and VLSI lab, Seow Miang Teo, Huan Qun Zheng, Wen-Sin Liew, Jun Tan, Chacko J. Deepu, Ding-Juan Chua, Wei Jie Eng, Yong-Fu Li, Xiaoyang Zhang, Mohammadreza Keshtkaran, Tianfang Niu, Lei Wang, Nankoo John , Wenfeng Zhao, Xiayun Liu, Zhang Zhe, Yibin Hong, David Tai Liang Wong, Rui Pan, for providing a nice working environment and technical helps and discussions. Special thanks to Dr. Mehran M. Izad for all great technical discussions and great times we had together during past 4 years and more importantly great friendship. I learnt a lot from him.

Most of all, I would like to express my deepest gratitude to my mother, my father, my brother, and my sisters for their tender love, guidance, and sacrifices. They are my greatest blessing in life. Lastly but most certainly not least, I would like to thank my wonderful wife, Azadeh, for her devoted love and unbounded support. She is the meaning of my life.

Table of Contents

Abstract

Mahmood Khayatzadeh
Department of Electrical and Computer Engineering
National University of Singapore
Prof. Yong Lian, Advisor

Power consumption arguably is the most important design factor in the recent millimeter-scale energy-autonomous systems like wearable and implantable biomedical sensor nodes. These systems have very stringent power requirement due to constraints on battery life and form factor. Therefore, ultra-low-power design is essential in the development of these nodes in addition to low cost. This work aims reducing the energy consumption of the digital subsection of a system-on-chip, particularly in two areas.

In the first part, this dissertation presents a new average-8T write/read decoupled (A8T-WRD) SRAM architecture for low power sub/near-threshold SRAM. The proposed architecture consists of several novel concepts in dealing with issues in sub/near-threshold SRAM including: (1) the differential and data-independent-leakage read port that facilitates a robust and fast read operation and alleviates issues in the half-selected cell (pseudo-write) while reducing the area compared to the conventional 8T cell; (2) the various configurations from 14T for a baseline cell to 6.5T for an area-efficient 16-bit cell. These configurations reduce the overall bitcell area and enable low operating voltage. Two memory blocks based on the proposed architecture at the size of 16 and 64 kb, respectively, are fabricated in 0.13 μm CMOS process. The 64 kb prototype has an active area of 0.512 mm^2 which is 16% less than that of the conventional 8T-cell-based design. In addition, this dissertation also pro-

poses a sizing technique to strengthen the write and access transistors based on the reverse narrow-width effect for the subthreshold SRAM in the advanced CMOS technologies. The technique is verified by a 16 kb SRAM chip in 65 nm technology. The measurement results show the chip consumes only 4.28 pJ/access in the best case with a supply voltage as low as 0.27 V. Based on the average of the measurements from 20 chips, the chip works from 30.8 kHz at 0.3 V while consuming 246 nW up to 2.42 MHz at 0.6 V while consuming 11.6 µW.

In the second part, this dissertation proposes an application-specific 16-bit microcontroller core which is customized for effective implementation of biomedical tasks. For the best energy efficiency, a subthreshold implementation of this core in a standard 0.13-µm CMOS process is presented. The measurement results show the microcontroller consumes only 2.62 pJ per instruction at 0.35 V and it is functional down to 0.22 V in the best case. On average it is capable of working from 52 kHz at 0.25 V to 6.1 MHz at 0.6 V. At the full speed, it consumes from 4.5 µW at 0.25 V to 90.2 µW at 0.6 V

Lastly, the proposed microcontroller and SRAM block in this work is implemented in a single-chip wireless ECG SoC. The entire SoC is capable of working at a single 0.7-V supply. At the best case, it consumes 17.4 µW in the heart rate detection mode and 74.8 µW in the raw data acquisition mode under a sampling rate of 500 Hz. This makes it one of the best ECG SoCs among the state-of-the-art biomedical chips.

List of Tables

List of Figures

List of Abbreviations

A8TWRD	Average-8T Write Read Decoupled
ADC	Analog to Digital Converter
ALU	Arithmetic Logic Unit
CMOS	Complementary Metal Oxide Semiconductor
CPI	Clock per Instruction
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
DRL	Driven Right Leg
DSP	Digital Signal Processor
DNM	Dynamic Noise Margin
DRV	Data Retention Voltage
DTMOS	Dynamic Threshold Metal Oxide Semiconductor
ECG	Electro Cardio Gram
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FSK	Frequency Shift Key
GBL	Global BitLine
GPIO	General Purpose Input Output
ISA	Instruction Set Architecture
LBL	Local BitLine
LOCOS	local Oxidation of Silicon
MAC	Multiply and Accumulate
MCU	Microcontroller Unit
MICS	Medical Implantation Communication Service
OOK	On Off Key
PMU	Power Management Unit
RBL	Read BitLine
RF	Radio Frequency
RFID	Radio Frequency Identification
RISC	Reduced Instruction Set Core
RNWE	Reverse Narrow Width Effect

RSCE	Reverse Short Channel Effect
SAR	Successive Approximation Register
SNM	Static Noise Margin
SoC	System on Chip
SPI	Serial Peripheral Interface
SRAM	Static Random Access Memory
STI	Shallow Trench Isolation
USART	Universal Synchronous/Asynchronous Serial Port
VLIW	Very Long Instruction Word
WBL	Write BitLine
WL	WordLine

Chapter 1
Introduction

1.1 Background

Energy autonomous systems, which can operate for very long time without the need for replacing or even using a battery, are one of the main directions of future electronic systems. Such systems can be used in variety of the applications like monitoring infrastructures e.g. bridges or buildings, biomedical wearable or implantable devices which monitor vital signals or stimulate nerves. The road towards such a system calls for the design of ultra-low-power circuits. In a typical energy autonomous system the required energy can be harvested from available sources like thermal, light, vibration, and so on. Table 1.1 summarizes the level of energy available for harvesting from common sources [1]. As can be seen from this table, considering the size of sensor nodes or implantable devices which is in the order of millimeters, the available power is in the order of μW or below. Therefore developing systems with this level of power consumption is crucial for these applications.

Table 1.1. Level of available energy for harvesting from common sources [1].

Source		Source Characteristics	Harvested Power
Photovoltaic	Indoor	0.1 mW/cm^2	10 µW/cm^2
	Outdoor	100 mW/cm^2	10 mW/cm^2
Vibration / Motion	Human	0.5 m @ 1 Hz 1 m/s^2 @ 50 Hz	4 µW/cm^2
	Industrial	1 m @ 5 Hz 10 m/s^2 @ 1 kHz	100 µW/cm^2
Thermal Energy	Human	20 mW/cm^2	30 µW/cm^2
	Industrial	100 mW/cm^2	1-10 mW/cm^2
RF	Cell Phone	0.3 µW/cm^2	0.1 µW/cm^2

Power reduction in an energy autonomous system should be performed in all hierarchical levels, i.e. choosing technology, circuit level, algorithm level, and so on, as well as all domains in each hierarchy level, i.e. digital, analog and RF. In this dissertation we mainly focus on the circuit level and partly on the architectural level of the digital subsection of a system. A digital subsection of a system-on-chip like a sensor node usually consists of some logics which can be in the form of a dedicated hardware or a programmable circuit like a microcontroller and some memory blocks like a static random access memory (SRAM) to store data. Both of these areas are investigated in this dissertation.

1.2 Sub/Near-Threshold Circuit Design

It is well-known that the best way to reduce power is to reduce supply voltage as it has quadratic impact on the power consumption. Operating at supply voltages as low as the threshold voltage of a transistor or even below it gains more attentions in the recent years, because it allows operating at the optimum minimum energy point and reduces power consumption up to orders of magnitude [2]. Therefore, various studies have been conducted on developing circuits in the sub/near-threshold regime.

Subthreshold operation provides substantial power reduction; however,

few concerns need to be carefully addressed in this regime. First of all, reducing the supply voltage considerably decreases the operating speed of the circuit. Therefore, the subthreshold technique is usually used for low-speed applications. Applying parallel processing to subthreshold circuits helps in improving the speed and enables applying this technique to medium-throughput applications.

The second important concern in subthreshold circuits is variation. Variation in the subthreshold regime is considerably higher than the conventional super-threshold regime. High variation in this region is caused by exponential dependency of the drain current to the threshold voltage in this region. Therefore any variation in the threshold voltage is translated exponentially to the variation in the drain current. High variation is a serious issue in the SRAM design in this regime as it requires very large number of cells, e.g. more than 6σ, working correctly to have acceptable yield. In addition, this high variation is also a concern in the timing closure of the digital circuits, especially for hold time.

Lastly, exponential behavior of the drain current in this region makes conventional transistor sizing and timing techniques in the conventional super-threshold regime less useful. As a result of the above concerns, subthreshold circuit design in general and subthreshold SRAM design in particular, is a challenging and interesting research which is still ongoing.

1.3 Research Objectives

The main objective of this work is to develop an ultra-low-power and low-voltage digital back-end for low-power system-on-chip applications like

wearable/implantable biomedical systems or sensor nodes. Two areas have been identified as bottle-necks in reducing total energy consumption of the digital back-end and the whole system. The first important block is static random access memory (SRAM) which is usually used to store captured signals. By increasing the complexity of the SoCs, larger SRAM blocks are required in the recent systems. This block needs to be always on and consumes considerable portion of the total power and the total area of the chip. The first part of this work will focus on the design of an ultra-low-power, low-voltage SRAM with acceptable dimensions.

The second critical block in the digital back-end is signal processing hardware. Biomedical sensor nodes, which are one of the main applications of this work, require intensive signal processing. This signal processing consumes considerable power during normal operation of the sensor node. In the second part, this work focuses on a programmable platform to perform biomedical signal processing with lower energy consumption compared to the available platforms.

1.4 Research Contributions

The contribution of this dissertation can be divided into two parts. The first contribution is proposing a new SRAM architecture, called average-8T write/read-decoupled cell, for reliable operation in the subthreshold regime. This architecture features differential-read, data-independent leakage, and pseudo-write tolerance along with dense area. Furthermore, a sizing technique based on the reverse narrow-width effect in the subthreshold SRAM is proposed which enables effective sizing of the transistors.

Second, novel microcontroller core architecture, customized for biomedical applications, has been proposed. This microcontroller has a few carefully selected DSP features like the multiply-accumulate instruction which enables efficient implementation of the computational-intensive tasks like filtering. Furthermore, a subthreshold implementation of this microcontroller core with a dynamic pipeline is provided which removes any hazard or stall in the pipeline and executes all instructions including interrupts in one clock cycle. This implementation also addresses hold violations due to large timing variations in the subthreshold regime.

A full ECG system-on-chip based on the above novelties has been designed and fabricated which shows the lowest energy consumption compared to the other recent works.

1.4.1 List of Publications

- M. Khayatzadeh, X. Zhang, J. Tan, W. Liew, Y. Lian, "A 0.7-V 17.4-µW 3-Lead Wireless ECG SoC", in *IEEE Trans. on Biomedical Circuits and Systems (TBioCAS),* (In Press).
- M. Khayatzadeh, Y. Lian, "An Average-8T Differential-Sensing Subthreshold SRAM with Bit Interleaving and 1k Bits per Bitline", in *IEEE Trans. on Very Large Scale Integration (VLSI) Systems,* (In Press).
- M. Khayatzadeh, X. Zhang, J. Tan, W. Liew, Y. Lian, "A 0.7-V 17.4-µW 3-Lead Wireless ECG SoC", in *IEEE Biomedical Circuits and Systems Conf. (BioCAS) Dig. Tech. Papers*, pp. 344-347, Nov. 2012.

1.5 Dissertation Overview

The rest of this dissertation is organized as follows. After a general introduction on the overall topic in Chapter 1, a review on the subthreshold designs is provided in Chapter 2. In this chapter, first a brief review on the SRAM design and various assist techniques is presented. Next, a detailed review on the previous subthreshold SRAM designs is provided followed by a section on the challenges of the subthreshold SRAM design. The second part of this chapter, describes a detailed review on the microcontroller designs aimed for ultra-low-power platforms and their subthreshold implementations.

Chapter 3 presents the proposed architecture for addressing issues in the design of a subthreshold SRAM. The various configurations of the proposed architecture are described and the potential of each configuration in reducing area or minimum operating voltage is explained. The proposed architecture is verified by two chip fabrications in 0.13 µm bulk CMOS process and their measurement results. In addition, utilizing reverse narrow-width effect in the design of a subthreshold SRAM is proposed at the end of this chapter. The idea is verified by another chip implementation in 65 nm bulk CMOS process.

Chapter 4 proposes a new 16-bit microcontroller architecture customized for biomedical applications. The proposed architecture is compared with the other commercial microcontrollers in terms of code size, number of clock cycle per task, and energy per task for 3 commonly-used biomedical tasks. A subthreshold implementation of the proposed microcontroller core is also presented. Finally, at the end of this chapter, a wireless ECG SoC consists of the proposed microcontroller in this chapter and the proposed SRAM block in

the previous chapter is demonstrated. In-vivo ECG recording on a volunteer in the lab using this system is performed successfully.

Finally Chapter 5 concludes the dissertation and proposes the next steps for continuing this work.

Chapter 2
Review of Subthreshold Circuit Designs

2.1 Overview

In this chapter existing subthreshold designs in the recent literatures have been reviewed. To limit the scope of this review and maintain the relevancy to this dissertation, this chapter focuses on two major designs. In the first part subthreshold SRAM designs are discussed. The review includes any SRAM design capable of working below 0.5 V. In the second part, subthreshold microcontrollers will be reviewed including implementations of the commercial microcontrollers in the subthreshold regime or the custom-designed ones.

2.2 Subthreshold SRAM Design Review

Ultra-low-voltage and low-power SRAM design is critical in the embedded systems such as biomedical implants, self-powered wireless sensors, and energy harvesting devices in which battery life or input power is of main concern [3-5]. Although ultra-low-voltage logic design has been well studied

and developed during past decades [6], SRAM design remains challenging and becomes more interesting due to rapid advancement of CMOS technologies and with increased demand of large memories in the embedded systems. By operating the SRAM in the sub/near-threshold regime, it is possible to reach the minimum energy point [2] which substantially reduces power consumption; however it adversely affects the speed and leads to high variation. Reduction in the speed maybe acceptable in some applications, but high variation is a serious problem in the SRAM, where a large number of cells have to work correctly on a chip. In addition, the weak on-current in the subthreshold regime could cause problem in the SRAM with large numbers of cells per bitline.

2.2.1 SRAM Basics

Fig. 2.1 depicts the conventional 6T and 8T SRAM cells which are widely used in the industry and academia. During a hold state the access transistors, $M_{AC1,2}$, are off and two cross-coupled inverters hold data. In a write operation, the access transistors are turned on and new data is placed on the bitlines (write bitlines in the 8T cell). Therefore, the discharged bitline overrides data in the cell.

In a 6T cell, the read operation is performed by pre-charging the bitline to Vdd and then turning on the access transistors. One side of the bitlines is

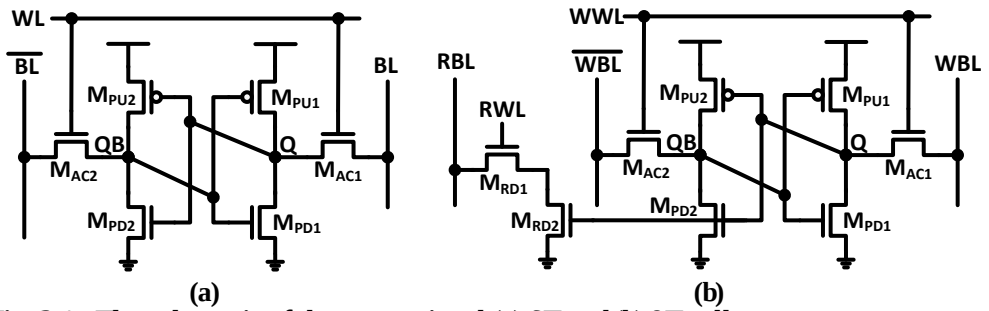


Fig. 2.1. The schematic of the conventional (a) 6T and (b) 8T cells.

discharged by data "0" in the cell and a sense amp detects the change on the bitlines and reads data. On the other hand, in an 8T cell the read stability is improved by decoupling the read port by adding two more transistors. In this cell, read stability is same as hold stability; however it provides a single-ended read port and needs more area. The 8T cell has also been used as a 2-port (1-read/1-write) register file.

Various metrics have been defined and used to measure stability of these cells during hold, read and write states. This subsection briefly describes each metric.

2.2.1.1 Butterfly curve

The butterfly curve is obtained by drawing the input-output characteristic of two cross-coupled inverters on a same graph, as shown in Fig. 2.2. In this graph one axis is the storage node Q and the other one is the storage node QB. The static noise margin is defined as the length of the largest square which can be fit in the smaller lobe of the butterfly curve. This curve is obtained by a DC analysis on a single cell and shows the static strength of the

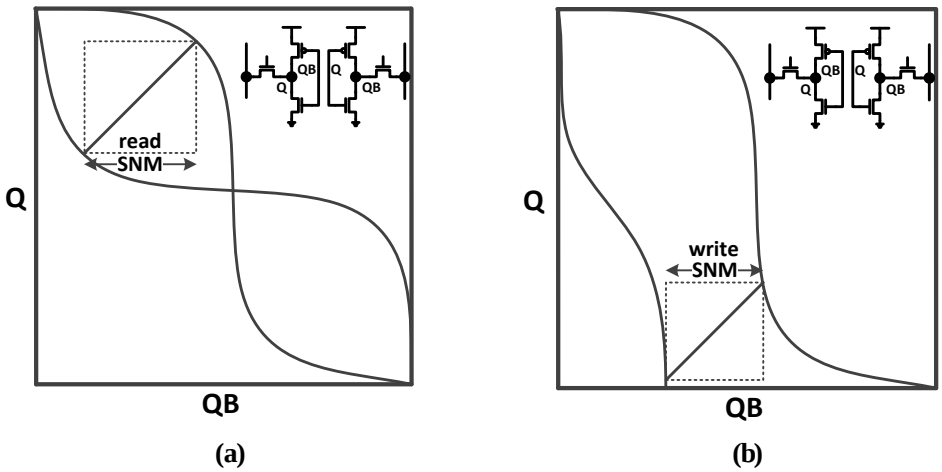


Fig. 2.2. Butterfly curve during a stable (a) hold or read state and (b) write state.

cell. This graph can be obtained for hold, read, and write states by applying the appropriate voltage to the wordlines and bitlines. For a stable write, this graph has only one stable point as shown in Fig. 2.2(b).

2.2.1.2 Bitline Sweep

This metric is used to calculate the write-strength during a write operation. In this analysis one or both bitlines (depending on the type of the cell and the write mechanism) are swept while the wordline is selected. The write margin is defined as the maximum bitline voltage which can perform a successful data-"0" write.

2.2.1.3 Wordline Sweep

This analysis is same as the bitline sweep analysis except that the bitlines are set to a predefined value and the wordline voltage is swept. The write margin is defined as the minimum wordline voltage (for an NMOS access transistor) which can perform a successful write.

2.2.1.4 N-curve

This method can be used for both read and write static stability analyses. In this analysis, a voltage source is connected to the side storing data "1" (for write noise margin analysis) and its value is swept from 0 to Vdd, as shown in Fig. 2.3. The current though this voltage source is captured as shown in this figure. The voltage distance between points B and C, and the maximum negative current between these two points are inversely proportional to the write strength.

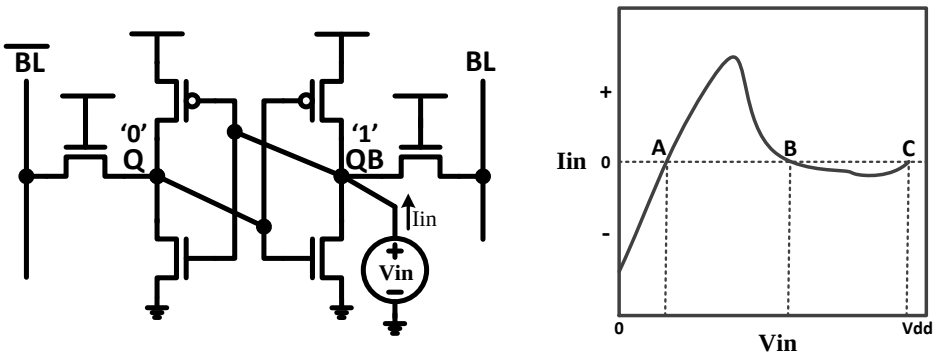


Fig. 2.3. (a) Simulation setup for the N-curve analysis. (b) Current waveform obtained from this analysis.

2.2.1.5 Dynamic Noise Margin

Increase in the parasitic capacitances and leakages in the advanced, deep-submicron technologies, requires analyses which includes the effect of these parameters from the adjacent cells, bitlines, and worlines in the design. In addition, larger variation in these advanced technologies causes more over-design and wider guard-band by using only the conventional static metrics. Dynamic analyses have been used to address above issues and get a more realistic behavior of the cell. Dynamic analyses are based on the transient simulation of the cell behavior. For example, the write time can be used as an indication of write strength. Another example is shown in Fig. 2.4. In this case, the value of the noise voltage, Vn, is changed and a transient analysis is

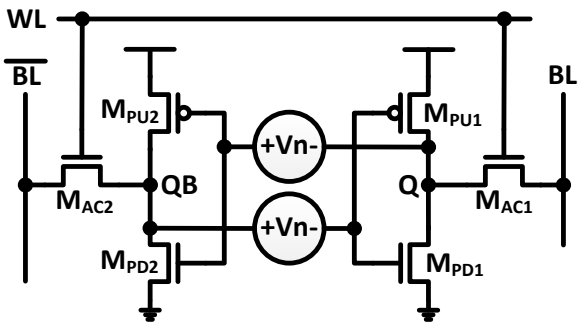


Fig. 2.4. An example of a simulation setup to calculate noise margin.

performed for each value to obtain the value at which the cell data flips. This value is considered as hold or read noise margin. As this is a transient analysis, the parasitic capacitances of the other cells and wiring must be added to the simulation to have a more realistic analysis.

2.2.2 SRAM Assist Techniques Review

To have a comprehensive review on the SRAM basics, various assist techniques which help improving the functionality of the memory in different operating modes should be reviewed. This subsection briefly reviews available techniques to improve read and write performance especially in the subthreshold regime.

2.2.2.1 Write Assist Techniques

Boosted Wordline: Boosting wordline voltage increases strength of the access transistors by increasing gate-source voltage and helps overriding cell data.

Negative-Boosted Bitline: Data is written in the cell by applying data 0 to the storage node via one of the bitlines. Considering an NMOS access transistor, applying negative voltage instead of zero to the bitlines helps the write-ability of the cell by making the access transistor stronger via increasing its gate-source voltage.

Supply/Ground Gating: Either disconnecting (or increasing) the ground terminal or disconnecting (or decreasing) the supply terminal of the cell improves the write-ability by reducing the strength of the cross-coupled inverters which holds old data.

2.2.2.2 Read Assist Techniques

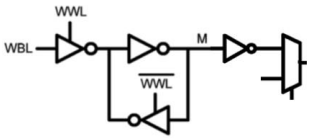
Tuning Wordline Voltage: Wordline voltage can be tuned to get the best read performance. In case of instability during a read operation, wordline voltage should be decreased to reduce the disturbance caused by the read operation. On the other hand, if developed voltage across the bitlines is not sufficient for correct sensing, wordline voltage can be increased (assuming that stability of the cell is still acceptable) to improve the discharge rate of the bitline by increasing gate-source voltage of the access transistor.

Boosted Supply: Gate voltage of the on NMOS in the cross-coupled inverters increases with increase in supply voltage. As a result, the read-ability improves by making discharge transistor stronger.

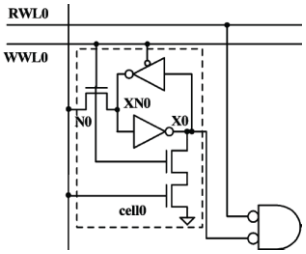
Negative-Boosted Ground: Like boosted supply, decreasing the ground voltage of the cell below zero, increases strength of the discharging NMOS by increasing its gate-source voltage.

2.2.3 Subthreshold SRAM Design

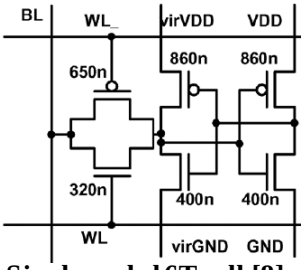
In early attempts towards the subthreshold SRAM, Wang and Chandrakasan [2] used a latch-based cell instead of the standard 6T block to implement a SRAM. In addition they used a mux-based read bitline to avoid leakage effect in the read path. In another work, Chen et al. [7] used a register-file-based cell with a modified read path to achieve an acceptable subthreshold operation. Both works obtained promising minimum operating voltage; however area overhead was very large. Therefore they are not being used anymore in the subthreshold regime.



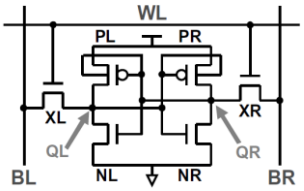
Latch-based cell with full-rail read and write [2].



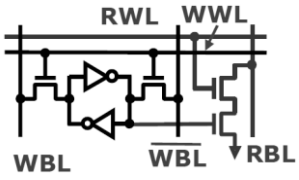
Register-file-based cell with a full-rail read and a single-ended write [7].



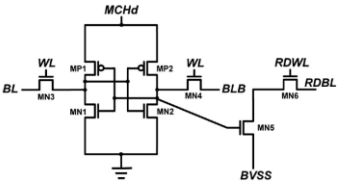
Single-ended 6T cell [8].



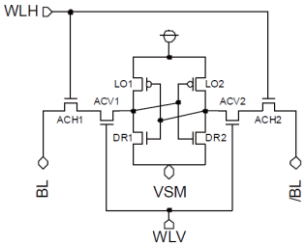
DTMOS 6T cell with differential read and write [9].



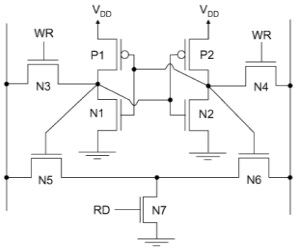
8T cell with a differential write and a single-ended read [10].



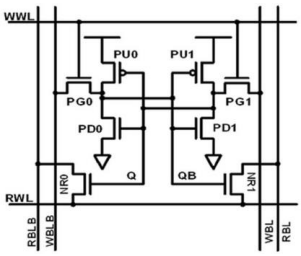
8T cell with a differential write and a single-ended read [11, 12].



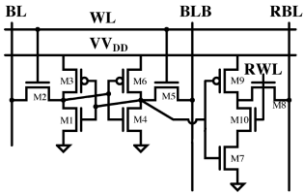
8T cross-point cell with differential read and write [13].



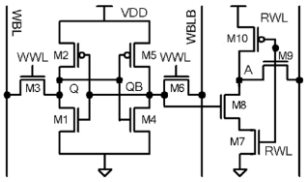
9T differential-read/write cell [14, 15].



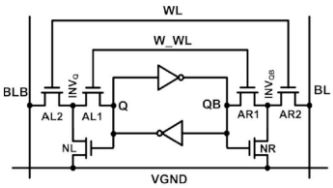
8T differential-read/write cell [16, 17].



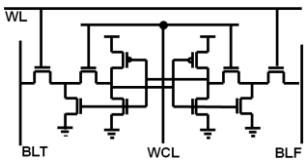
10T cell with a single-ended read and a differential write [18].



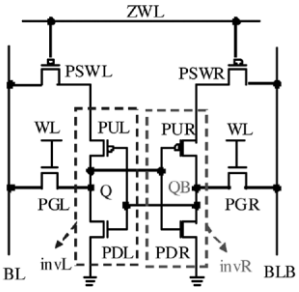
10T cell with a single-ended read and a differential write [19].



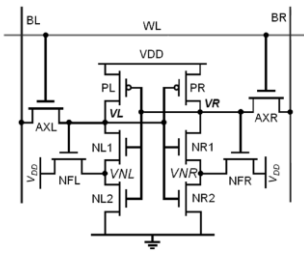
10T differential-read/write cell [20].



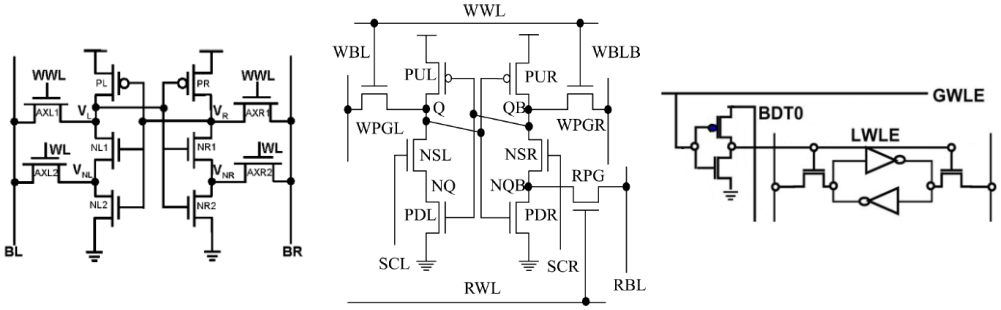
10T differential-read/write cell [21].



8T D2AP differential-read/write cell [22].



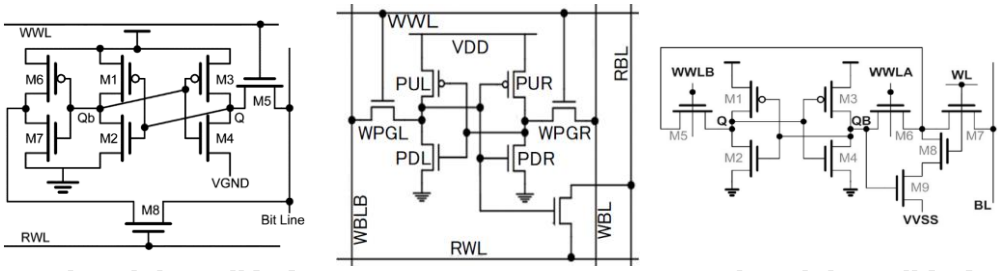
10T Schmitt-Trigger-based cell with differential read and write [23].



10T Schmitt-Trigger-based cell with differential read and write [24].

9T cell with a differential write and a single-ended read [25].

Column-decoupled 8T cell with differential read and write [26, 27].



Single-ended 8T cell [28].

7T cell with a differential write and a single-ended read [29].

Single-ended 9T cell [30].

Fig. 2.5. Overview of the previous subthreshold SRAM cells in the recent literatures.

Zhai et al. proposed a single-ended 6T cell which used gated supply and ground to assist the write operation and upsized transistors to overcome variations [8]. A transmission gate is used as the read/write port. This transmission gate allows driving the bitline rail-to-rail and eliminates the need for a sense amplifier. It should be noted that, this is valid only for a very short bitline otherwise the large capacitance of the bitline will reduce cell stability. A hierarchical bitline is required for a large number of cells.

Hwang and Roy in [9] proposed using the dynamic threshold voltage MOS (DTMOS) circuit in the cell by tying the body of the PMOSes to their gate. This technique dynamically changes the threshold of the PMOSes by changing the body voltage. The proposed technique is only applicable to the subthreshold regime so as to prevent turning the body diode on. Furthermore, it requires separate wells for two PMOSes which causes large area overhead.

The 8T cell first introduced by Chang et al. in [10] to mitigate the effect of variation in the 32 nm node and beyond. Later this architecture and its variations have been successfully used for the sub/near-threshold operation [11, 12, 31, 32]. This architecture decouples the read and write ports which allows independent improvement of the read and write noise margins. On the other hand, it suffers from the single-ended read port, data-dependent leakage of the read port, and instability of the half-selected cells during a write. Verma and Chandrakasan in [11] used the 8T cell to implement an SRAM in the subthreshold. They connected the read port of all cells in the same row to a virtual ground instead of a ground. A buffer controls voltage of this virtual ground. However, boosted voltage is required in the subthreshold to provide a strong buffer for this virtual ground. This virtual ground alleviates the data-dependent-leakage problem. Furthermore, they reduced supply voltage of the cell to assist the write operation. It should be noted that stability of the cells sharing a same supply should be addressed carefully. To avoid the pseudo-write problem in the half-selected cells, they separated the wordlines of the cells in a same row, which adds area overhead. The sense-amplifier redundancy concept has been introduced in their work to improve read performance in the subthreshold regime. Sinangil et al. applied configurability to the above mentioned techniques in [12] and designed an ultra-dynamic voltage scalable SRAM which works from 250 mV up to 1.2 V. Kim et al. in [32] used the conventional 8T cell and tried to compensate leakage by adding a programmable pull up and a replica bitline. In addition, to mitigate the effect of variation in single-ended reading, they made the trip point of the sense amplifier programmable. The reverse short-channel effect was used in this work to make

the transistors stronger in the subthreshold.

Yabuuchi et al. in [13] presented a cross-point 8T cell which allows pin-pointing a cell by selecting the rows and columns. In addition they used the negative-boosted ground and the negative-boosted bitline techniques to assist the read and write operations, respectively. Using this method the read and write operations can be optimized independently. This technique requires negative voltage. Furthermore, stability of the half-selected cells during the write should be addressed carefully.

A 9T differential-read cell has been proposed by Liu and Kursun in [14]. This cell is same as the standard 8T cell with read decoupling at both sides. The number of transistors is reduced by sharing the access transistor between both sides. This configuration allows a differential read which helps improving read performance in the presence of high variation. This bit cell still suffers from the pseudo-write issue in the half-selected cells during a write operation. In addition, data-dependent leakage should be addressed carefully. Lutkemeier et al. used an optimized version of this cell with multi-V_{TH} transistors in their system as reported in [15]. To improve this cell, Wu et al. in [16] and Kulkarni et al. in [17] proposed to share the read access transistor among the bit cells in the same row and eliminate it from the bit cell. This gives an 8T differential-read bit cell which improves area over the 9T cell; however it still suffers from the data-dependent leakage and pseudo-write issues. The pseudo-write issue can be addressed separately e.g. during floor planning of the array by breaking the wordline in each row or by writing back data of the unselected columns.

In an attempt towards making leakage of the read port independent from stored data in the cell, two 10T cells proposed in [18] by Calhoun and Chandrakasan and [19] by Kim et al. In these works, two additional transistors are added in the read decoupling path to reduce leakage of the unselected rows in each column. [18] ensures that leakage is minimum at all times by either having two off transistors between the bitline and ground or having the same voltage across the drain and source of the access transistor of the unselected cells. This is performed by pulling up the other side of the access transistor which is not connected to the bitline. The work in [19] tries to reduce read-port leakage only by applying the same voltage across the access transistor.

Chang et al. in [20] proposed a 10T cell which allows a differential read and solves the pseudo-write problem for the half-selected cells at the same time. However, due to the presence of feedback in the cell and two series access transistors in the write path, write strength is low and it requires virtual ground and boosted wordline voltage for a reliable write operation. A similar cell has been proposed by Clerc et al. in [21] in which the gate of the read-decoupling transistor is connected to the opposite side. This connection removes the internal feedback of the cell and eases the write process.

A differential data-aware power-supplied (D^2AP) has been proposed in [22] by Chang et al. in order to improve minimum operating voltage by expanding write and read stability. In this work, instead of connecting the source of the PMOSes in the cross-coupled inverters to the supply, they are connected to the bitlines via two PMOSes. In this architecture, during a write, the supply of the corresponding inverter is collapsed, while the other inverter is powered on. This mechanism provides a fast write; however, the stability of the other

cells in the same column is affected and the duration of the write should be controlled carefully. The same concern rises during a read operation. Therefore accurate timing is required for this cell which makes its application in the sub/near-threshold difficult, considering high variation in this regime.

[23, 24] proposed two Schmitt-Trigger-based 10T cells in the 0.13 µm process. The feedback nature of the Schmitt-Trigger helps stability of the cell by modulating the switching threshold of the inverter; however it adds considerable area overhead. These cells show very impressive minimum operating voltage. These cells provide a differential read port which is desirable in the ultra-low voltage applications.

A 9T cell has been proposed in [25] by Chang et al. Like the 8T read-decoupled cell, this cell has two separate paths for the read and write. However, two additional transistors are added in series with the transistors of the cross-coupled inverters. During a write, depending on data, one of these transistors is turned off to break the feedback loop for a more reliable write. This scheme allows a cross-point-selective assist for the write. During a read operation, both of these series transistors are turned off to decouple the storage node from the read bitline. However, this scheme causes the storage node to be float, therefore careful timing is required to prevent data loss during a read. In addition the column half-selected cells and data-dependent leakage of the read path should be addressed carefully to have a reliable operation. This work achieves very impressive minimum operating voltage, i.e. 130 mV.

Joshi et al. in [26] and Anh-Tuan et al. in [27] proposed a column-decoupled 8T cell to solve the disturbance issue in the half-selected cells. In

this architecture, the wordline-select pulse of the conventional 6T cell was connected to a local driver. The supply of this local driver comes from a vertical selection control. This configuration allows cross-point selection.

A single-ended 8T cell with a few assist techniques proposed in [28] by Tu et al. to reduce active power due to discharging the highly-capacitive bitlines. Despite the differential read, in the single-ended architecture the bitline is discharged only if data to be read is zero. However the single-ended read has less noise immunity especially in the presence of high variation in the subthreshold. Furthermore, the write in this cell is also single-ended which needs assist circuits for a robust operation. Virtual ground was used to reduce strength of the inverter at the side which is going to be written.

A 7T L-shape cell was proposed by Chen et al. in [29]. They used only one transistor for read decoupling instead of two in the conventional 8T cell and selection was performed by controlling the source of this read decoupling transistor. In addition the boosted read-bitline, asymmetric-V_{TH}, and offset cell-VDD biasing techniques were used to improve write performance. It should be noted that, changing forward and reverse V_{TH} in the proposed technique requires hot-carrier injection after the fabrication. In this work, the pseudo-write issue is addressed by adding the write-back scheme during a write operation.

Tu et al. in [30] presented a single-ended 9T cell which features a cross-point data-aware write. Therefore it has no pseudo-write problem. On the other hand, because it has a single-ended architecture it requires assist circuits for a reliable write operation. The negative bitline-boosting technique was

used in this work to improve writeability. Furthermore, due to the single-ended read port, sensitivity of the read operation should be addressed carefully. In this work an adaptive read operation timing tracing (AROTT) circuit was proposed to track PVT variation.

2.2.4 Subthreshold SRAM Design Challenges

In the sub/near-threshold regime, the conventional 6-transistor (6T) SRAM shows poor functionality [33], especially on read stability, write ability, and half-selected cells. Techniques and methods have been proposed to address these three issues in the past [8-12, 14, 16, 17, 19, 20, 25, 28, 32]. To address the issues related to read stability, decoupling the read port from the write has been introduced [10] to separate the read and write paths, which allows each operation to be optimized individually at the cost of additional transistors, such as 8T [11, 12, 16, 17, 28, 32], 9T [14], and 10T [19, 20, 33] cells. Apart from the read decoupling technique, other methods have been reported recently such as upsizing the standard 6T [8], DTMOS cell [9], dynamic read decoupling [25], and the Schmitt-Trigger-based cell [24]. As most of the conventional read decoupling techniques use single-ended read schemes, detection threshold varies considerably due to leakage. For example, in an 8T cell, leakage of each cell depends on data stored in the cell. This dependency causes voltage variation on the read bitline, i.e. read bitline level changes with the number of cells storing data 0 or 1 on the same bitline. Such variation makes the detection more challenging. Leakage compensation using the replica bitline [32], additional transistors (10T cell) to make leakage independent of data [19, 20, 33], virtual ground [11, 12], and negative wordline voltage [25] are some of the techniques which have been proposed to combat

the effect of data-dependent leakage at the cost of more area and/or power than that of the conventional 8T cell. To overcome difficulties in the single-ended read, several techniques were proposed to maintain a differential-read port by adding more transistors (10T [20, 24], 9T [14]) at the cost of extra area or using virtual ground/supply for each row [16, 17] which limits the number of cells per wordline, decreases stability of the half-selected cells and increases area and power. Therefore, a differential-read SRAM with a robust read scheme and minimum area is still an unsolved problem. Data-independent leakage is one of the key aspects in such a read-robust cell.

To improve writeability, either the cell being written should be made weaker or the access devices should be made stronger. Two methods were proposed to weaken a cell, i.e. reducing supply/increasing ground voltage [8, 11, 12, 22, 28], and breaking the feedback loop of the cell [25]. Both of these methods affect stability of the half-selected cells, which will be discussed later, and may require extra area for either additional power/ground rail or a feedback-breaking device. On the other hand, upsizing access transistors [14, 16, 17], boosting wordline voltage [11, 20], or using reverse short channel effect [19, 32] were reported to strengthen the access devices. The effect of upsizing is diminished in the deep subthreshold and so does reverse short channel effect in the super-threshold. Boosting wordline voltage needs additional charge pump circuit to boost the voltage.

To maintain stability of the half-selected cells, i.e. the cells which are selected during a write operation on a shared wordline but not being written (pseudo-write problem), the simplest way is to prevent the wordline from sharing among the cells in the same row [11, 17, 25] at the cost of extra area.

Two other ways are possible: (1) to perform a write-back operation [16, 19], i.e. reading and writing back its own data on the cells which are not going to be written at the expense of speed and power; (2) to allow cross-point selection e.g. by adding two more transistors, i.e. 10T cell, in [20] that avoids the problem of half selection at the cost of 61% area penalty with respect to the 8T cell. Compared to the separate wordline, the cross-point selection 10T cell allows wordline sharing among more cells, i.e. high level area saving, but adds extra area in the cell level. The amount of area saving/overhead of this method with respect to a separate wordline for each word depends on the technology, memory macro size and minimum operating voltage and should be examined during the design. As a result, a pseudo-write-tolerable architecture with acceptable size, e.g. same as the 8T cell, is required to reduce power while saving area.

2.3 Subthreshold Microcontroller Designs Review

An extensive analysis over 21 microarchitectures has been performed by Nazhandali et al. in [34] and in more details by Zhai et al. in [35] to evaluate the effect of the microarchitecture on energy efficiency of sensor network processors. In this analysis the best energy efficiency obtained from a very simple 8-bit architecture with a compact instruction set. However, this architecture causes a considerable performance penalty. On the other hand, a 16-bit or 32-bit design with a Harvard architecture and a 2-stage pipeline gave the best balance between energy and performance. Furthermore, a Harvard memory architecture provided better energy consumption with respect to the Von Neumann architecture. Another important observation in this work was that a more complex instruction set which leads to compact code size out-

weighs extra complexity of the control logic. As a result a more complete instruction set is desirable to reduce the size of the memory required to store the code and the number of accesses to the code memory. In their work, they proposed an 8-bit subthreshold processor with a 2-stage pipeline which achieved 2.6 pJ/inst. at 360 mV and 833 kHz [35, 36]. They used a reduced set of the standard cells to implement the processor and a 2 kb mux-based array [2] as the microcontroller memory. They have implemented another variant of their processor in [37] in which they achieved 3.5 pJ/inst. at 350 mV. They used body biasing to mitigate the effect of variations in the subthreshold. In their work, they concluded that in the subthreshold regime the energy consumption is almost independent of threshold voltage. Finally this group proposed usage of chip multi-processor in [38] to leverage the effect of low performance in the subthreshold regime. In their system multiple slow cores working at a subthreshold supply, were clustered with a shared cache which operated faster. Therefore each block worked at its optimal speed.

Seok et al. in [39] proposed a subthreshold sensor platform based on a subthreshold custom processor named Phoenix which consumed 2.8 pJ/cycle at 0.5 V supply voltage and 106 kHz speed. Extensive power gating was used in their work to obtain low standby power of 30 pW. The Phoenix system was an event-driven system which was waked up with a very low power timer to perform its tasks.

A subthreshold implementation of the commercial PICmicro microcontroller [40] has been reported by Jocke et al. in [41]. They achieved the minimum energy consumption of 1.51 pJ/inst. at 280 mV with operating speed of 475 kHz. They implemented an ECG system-on-chip based on this microcon-

troller. However, the implemented memory in this work was operating at normal high voltage. As a result this chip required two different supply voltage and level-shifters at the interfaces.

Subthreshold implementation has not only being used for low speed applications. Kaul et al. have reported a motion estimation accelerator which is capable of working from 4.3 MHz (at 230 mV) to 2.4 GHz (at 1.4 V) [42]. They achieved the minimum energy consumption at 320 mV with performance of 411 GOPs/W at 23 MHz. In another work a subthreshold multi-standard JPEG co-processor is reported in [43] by Pu et al. A configurable V_{TH}-balancer was proposed in this work to compensate variations in the subthreshold regime. Furthermore, parallel processing was used to achieve required throughput for a real-time operation. As another example, Seok et al. in [44] reported a 1024-point Complex FFT core which was working at 0.27 V at the speed of 30 MHz while consuming 17.7 nJ/transform. Pipelining and parallel processing were implemented in this work to improve the speed.

A subthreshold microcontroller based on the commercial MSP430 microcontroller [45] with an integrated subthreshold SRAM has been reported in [46] by Kwong et al. They also integrated a DC-DC converter to generate subthreshold voltage. They achieved energy consumption of 27.2 pJ/inst. for the whole system and around 7 pJ/inst. for the microcontroller core only at 0.5 V. A subthreshold standard cell library and a variation-aware timing methodology were used to develop this design. The integrated subthreshold SRAM was an important aspect of this design which allowed a single low supply for the whole system and improved power consumption of the whole system. In another work by these authors [47], they integrated this microcontroller with a

few accelerator engines like FIR, FFT, median, and CORDIC to get a low power platform for biomedical signal processing. The reported platform was functional down to 0.5 V and achieved considerable energy reduction compared to the CPU-only systems. Another subthreshold implementation of the MSP430 core has been presented in [48] by Bol et al. An on-chip DC-DC converter, controlled by an adaptive voltage scaling system, was employed in this work to ensure a reliable 25-MHz operation at all conditions. The core was functional down to 0.4 V and 25 MHz; however, the SRAM was working at full voltage i.e. 1 V.

Chen et al. [3] proposed a full system, with a solar cell, battery, and power management, based on an ARM Cortex-M3 [49] microcontroller which was operating at near-threshold voltage. It achieved 73 kHz speed at 0.4 V while consuming 28.9 pJ. A low-power near-threshold SRAM was also designed and integrated into this system. Another implementation of the ARM Cortex-M3 microcontroller with an integrated memory was reported in [5] by Sridhara et al. This implementation operated at the speed of 7 kHz at 0.5 V while consuming 29 nW/kHz.

A reconfigurable fabric in 32 nm CMOS process was proposed by Agarwal et al. [50] which was capable of working from the subthreshold up to a normal supply. This design can be considered as a small subthreshold FPGA which can be used as an accelerator engine.

Ashouei et al. reported [51] a subthreshold biomedical signal processor based on a new DSP engine, named CoolFlux [52], from NXP. They achieved 13pJ/cycle at 0.4V and 1 MHz. However, they used commercial memory

which was not working at the subthreshold regime. As a result they needed various supply voltages and had to insert level shifters at all core-memory interfaces. Another DSP for mobile applications was proposed by Gammie et al. [53] which was based on a VLIW DSP from TI. The core was fabricated in the 28 nm CMOS process and it was operational down to 3.6 MHz at 0.34 V where it consumed 720 µW. A new statistical static timing analysis was proposed in this work to accurately model V_{TH} variation in low supply voltage and to prevent over-design.

A custom-designed 32-bit microprocessor has been reported in [54] which was consuming 10 pJ/cycle at 0.54 V and 540 kHz. The core was developed at STMicroelectronics and named Reduced Energy Instruction Set Computer (ReISC). An 8T memory was integrated into this system which was working at the same voltage as the core.

Finally, a 32-bit subthreshold processor based on a custom-designed configurable VLIW core, named CoreVA, has been reported by Lutkemeier et al. in [15, 55]. The core featured Harvard architecture with a 6-stage pipeline and was implemented using a subthreshold standard cell library. A 9T subthreshold SRAM was also developed for this system. This processor achieved energy consumption of 9.94 pJ/cycle at 325 mV and 133 kHz.

Table 2.1 summarizes the specifications of these recent subthreshold processor designs.

Table 2.1. Comparison of the recent works in the subthreshold microcontroller and microprocessor.

Reference	Architecture	Speed Range	Memory Supply Domain	Application	Special Features
[35-37]	8-bit Custom-Design	Low	Same as logic	Wireless Sensor Nodes	Body-Biasing
[39]	8-bit Custom-Design	Low	Same as logic	Wireless Sensor Nodes	Event-Driven System
[41]	8-bit PICmicro	Low	High Supply	Biomedical	
[42]	Motion estimation accelerator	High	N.A.	Mobile	Parallel Processing
[43]	JPEG co-processor	Med-High	N.A.	Mobile	Vth-balancer Parallel processing
[44]	FFT co-processor	High	N.A.	Low Power Processing	Parallel Processing
[46,47]	16-bit MSP430	Low	Same as logic	Biomedical	On-Chip DC-DC HW Accelerator
[48]	16-bit MSP430	Med	High Supply	Wireless Sensor Nodes	On-Chip DC-DC Adaptive Volt. Scaling
[3]	32-bit ARM Cortex-M3	Low	Same as logic	Wireless Sensor Nodes	Solar cell, Battery, Power Management
[51]	24-bit CoolFlux	Low	High Supply	Biomedical	
[53]	32-bit TMS320C64x+	Med	Same as logic	Mobile	
[54]	32-bit, ReISC Custom-Design	Low	Same as logic	Low Power Processing	
[15, 55]	32-bit, CoreVA Custom-Design	Low	Same as logic	Low Power Processing	

Chapter 3
Ultra-low-power, low-Voltage SRAM Design

3.1 Overview

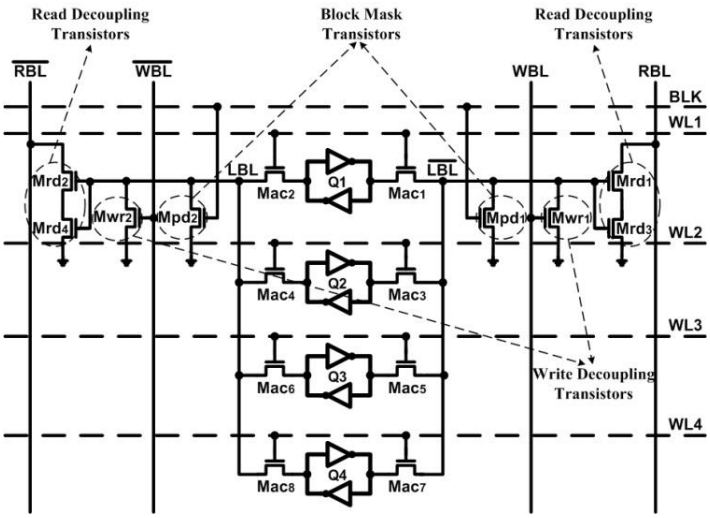
In this chapter, a new average-8T write/read decoupled architecture for ultra-low-voltage SRAMs is proposed with a number of circuit techniques to address the issues mentioned in Section 2.2.4 while maintaining minimum area. The proposed structure allows a differential read with data-independent leakage on the read port, a pseudo-write-tolerable and ultra-short-segmented write to improve the write operation with smaller "per cell area" than previously reported 8T-cell in the same technology. In addition its various configurations allow minimum operating voltage versus area optimization. Two memory blocks based on the proposed architecture at the size of 16 and 64 kb, were fabricated in the 0.13 µm CMOS process. In addition, we propose for the first time to utilize reverse narrow-width effect (RNWE) as an efficient sizing technique

in the subthreshold SRAM design for performance enhancement. The technique is verified in another 16 kb SRAM block in the 65 nm technology based on the proposed average-8T architecture. Measurement results were presented for three memory blocks. The rest of this chapter is organized as follows. In Section 3.2, the proposed average-8T SRAM architecture is introduced with detailed explanations on how the proposed techniques address the existing issues. Section 3.3 provides a detailed analysis on various configurations of the proposed architecture. RNWE-aware sizing is presented in Section 3.4. Chip implementation details and measurement results of the sample designs based on the proposed architecture and RNWE-aware sizing are demonstrated in Section 3.5. Section 3.6 summarizes this chapter.

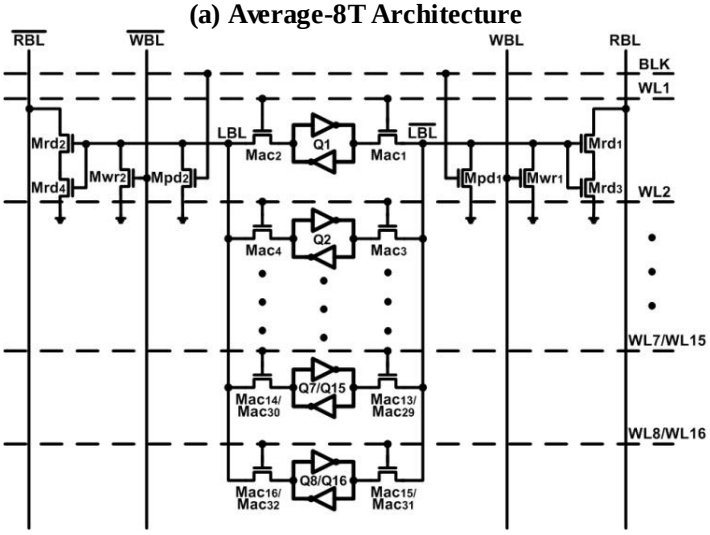
3.2 Average-8T Write/Read Decoupled (A-8T-WRD) SRAM Architecture

The basic architecture of the proposed write/read decoupled SRAM block is illustrated in Fig. 3.1(a) and its various configurations are given in Fig. 3.1(b)-(d). The number of transistors in each block depends on the number of bits stored in the block. It varies from storing 1 bit to 16 bits per block. This architecture fills the gap between the 8/10T-cells and the conventional 6T cell. The selection for a specific configuration is based on the minimum operation voltage and area constraint, which will be discussed in Section 3.3. In the following, we use the average-8T block, as shown in Fig. 3.1(a), to illustrate the operation of the proposed SRAM.

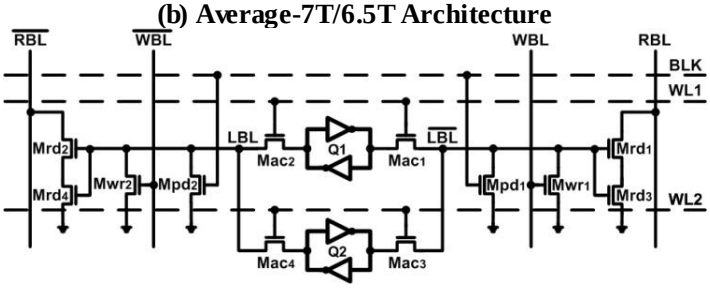
The average-8T block holds 4 bits through 4 back-to-back connected inverters. The storage nodes are connected to the local bitlines (LBL and $\overline{\text{LBL}}$)



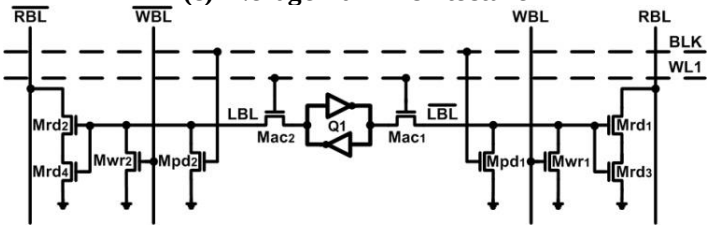
(a) Average-8T Architecture



(b) Average-7T/6.5T Architecture



(c) Average-10T Architecture



(d) 14T Architecture

Fig. 3.1. (a) Basic architecture of the proposed block which stores 4 bits. (b)-(d) Its various configurations which store 16, 8, 2, and 1 bits, respectively.

via $Mac_{1\text{-}8}$. These local bitlines are decoupled from the write bitlines (WBL and $\overline{WBL}$) via $Mwr_{1\text{-}2}$ during a write operation and from the global read bitlines (RBL and $\overline{RBL}$) via $Mrd_{1\text{-}4}$ during a read operation. This new write/read-decoupled (WRD) technique allows complete isolation of these 4 bits. In an idle state, when neither a write nor a read operation occurs on a block, the local bitlines are float; therefore they may turn $Mrd_{1\text{-}4}$ on and interfere with the read operation of other blocks or increase standby leakage. Block mask transistors (Mpd_1 and Mpd_2), as shown in Fig. 3.2(a), are introduced to assure that LBL and $\overline{LBL}$ are low when the block is not selected and their leakage is minimized.

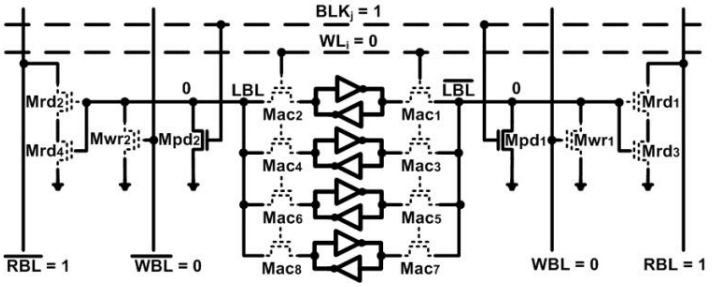
A distinct difference between the proposed technique and the conventional 8T cell with a simple hierarchical bitline [31] or high-speed 6T register files with a hierarchical bitline is data-independent leakage in the read bitline which is achieved by newly introduced block mask transistors. Data-independent leakage is very desirable in the subthreshold regime in which I_{on}/I_{off} ratio is very low. The second difference is the write decoupling transistors which allow complete isolation of the bits in the cell and minimize the write disturbance in the half-selected cells. The Last difference when comparing to the hierarchical 8T cell [31] is the differential read port in the proposed architecture. The area overhead of this differential decoupling is reduced by sharing the decoupling transistors among multiple bits.

3.2.1 Read Operation

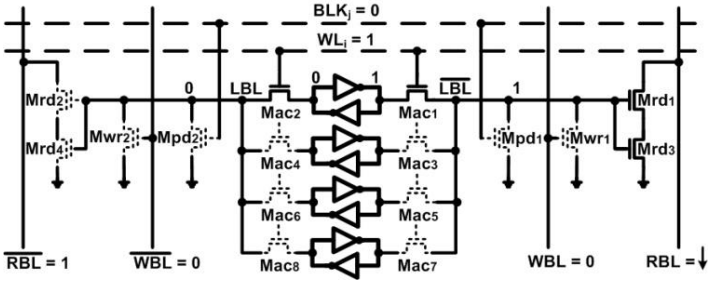
A new read decoupling technique that decouples more than one bit at both Q and QB sides ($Mrd_{1\text{-}4}$ in Fig. 3.1(a)) is proposed to improve read ro-

bustness. The access transistor of the traditional 6T cell is used as selection switch among data bits. This technique provides three interesting features. First, the leakage is no longer data-dependent. According to Fig. 3.2(b), during a read operation, Mpd_1 and Mpd_2 of the intended block are turned off and the access transistors of the intended bit (e.g. $Mac_{1,2}$) are turned on. The stored data turns one of the read-decoupling pair transistors ($Mrd_{1\text{-}4}$) on and discharges one of the pre-charged read bitlines (RBL or $\overline{RBL}$). In other unselected blocks on the same bitline, both local bitlines are pulled down via $Mpd_{1,2}$, therefore the $Mrd_{1\text{-}4}$ of the unintended blocks are off and their leakage is not only minimized but also independent of data stored in that block.

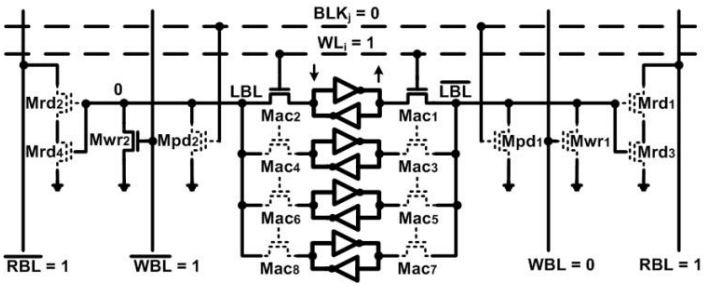
In the technology we used, NMOS transistors are much stronger than PMOS transistors especially in the subthreshold regime. As a result NMOS transistors are used in the read port ($Mrd_{1\text{-}4}$) to have faster discharge in the high-capacitance global read bitlines (RBL or $\overline{RBL}$). Using this scheme, the local bitlines have to be pre-discharged to ground. Two stacked transistors in the read decoupling path are used to further reduce the leakage of the unselected block. Another option is to use virtual ground instead of two stacked transistors. However, the area required for additional ground rail and power consumption of virtual ground buffer should be taken into account. In the technology we used, the number of available metal layers limits the number of horizontal and vertical rails resulting larger cell size for virtual ground approach. Second, this data-independent minimum-leakage feature facilitates implementation of large number of bits per bitline, e.g. 1024 in our design example, which yields denser memory design. Third, it provides a differential read port which allows more robust sense amplifier design especially in the



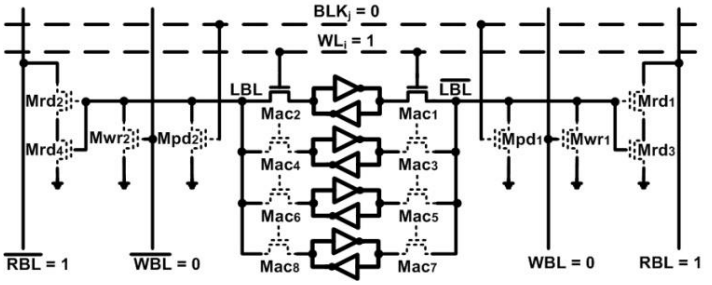
(a) Unselected / Idle state



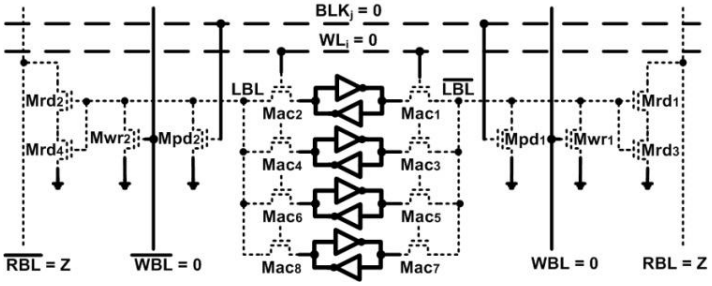
(b) Read state (reading "0")



(c) Write state (writing "0")



(d) Half-selected blocks during write (in the same row)



(e) Minimum leakage state

Fig. 3.2. Operation of the proposed block at various states.

presence of high variations in deep sub-micron technologies and ultra-low voltages. During a read operation high data is transferred via an NMOS transistor to the gate of the read decoupling transistors. NMOS transistors are chosen in the access path to help the single-ended write process which is described in the next section. To mitigate the effect of this NMOS on the read performance, the gate voltage of the NMOS (wordline voltage) is boosted. This boosting is also helpful during the write operation to improve the write strength of the block. Simulations show that a 50 mV boost is sufficient to allow correct operation in the sub/near-threshold regime i.e. supply voltage range from 0.25 V to 0.5 V. In addition to the wordline voltage boost, reverse short channel effect (RSCE) [19] is used to improve the driving strength of the access and pull down transistors.

Compared to the conventional 8T read-decoupled cell, the proposed block slightly degrades the read noise margin due to the additional cells connected to the same bitline and boosted wordline voltage; however Monte Carlo simulation, taking into consideration the effect of other cells, shows acceptable robustness and less than 9% reduction in read noise margin at 300mV. To further investigate read stability, which is useful for stability of the half-selected blocks during a write too, dynamic noise margin has been simulated. To consider all side effects in this simulation, the whole block, as shown in Fig. 3.3, has been simulated. In addition, all wiring parasitic capacitance were obtained from post-layout extraction and has been added to the circuit. As rise-time and fall-time of the wordline (WL) and block select (BLK) signals are important in the stability simulation, these values are calculated from a post-layout simulation of the whole address decoding logic. Then multiple

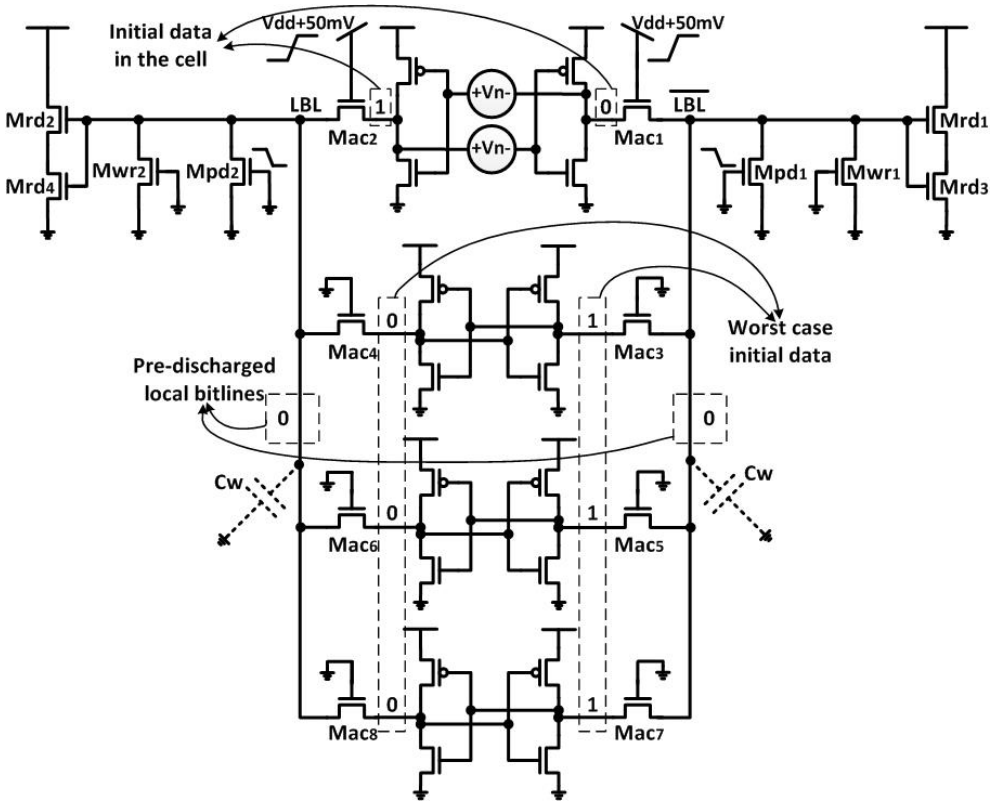


Fig. 3.3. Simulation setup for the dynamic/static noise margin analysis considering all side effects. In this setup the capacitance of LBL to the capacitance of the storage node is 2.7fF to 1fF.

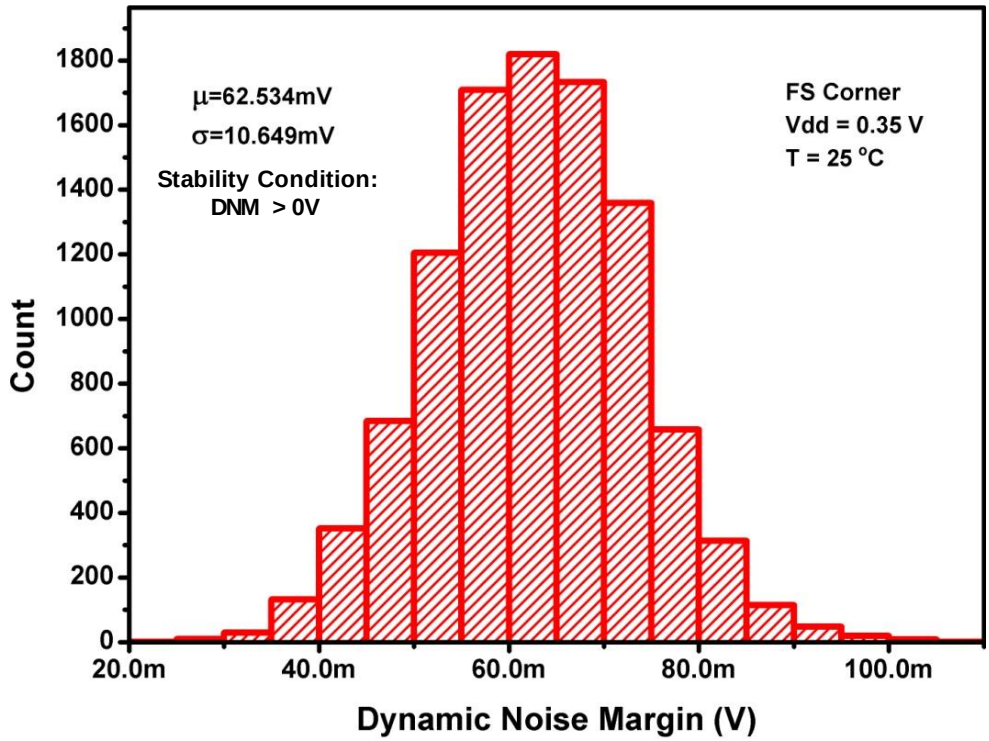


Fig. 3.4. Monte Carlo simulation result of dynamic noise margin analysis of read / pseudo-write stability at the worst case corner based on the transient analysis on the circuit shown in Fig. 3.3. The dynamic noise margin is considered as the noise voltage value, i.e. Vn, at which cell data is toggled.

transient analyses with various noise voltage values have been performed on this circuit. Dynamic noise margin is considered as the noise voltage value at which cell data is toggled. Fig. 3.4 shows the result of Monte Carlo analysis for dynamic noise margin at the Fast-Slow corner which is the worst corner. As can be seen from Fig. 3.4, the block is stable with acceptable margin at the worst case corner.

It should be noted that at the start of a read operation the block mask (pull down) transistors should be turned off before turning on the access transistors to prevent data loss at the storage node. At the end of the procedure the access transistors should be turned off first. Therefore BLK and WL signals should be carefully designed to have non-overlapped switching. Fig. 3.5(a) shows the timing diagram of a read operation.

3.2.2 Write Operation

A write operation, as shown in Fig. 3.5(b), is performed by selecting the intended block and bit in the same way as a read. Depending on written data, one of the write decoupling transistors, Mwr_1 or Mwr_2, is turned on and pulls down the storage node in the intended bit (Fig. 3.2(c)). Boosted wordline voltage (by 50 mV as described in read improvement subsection) and longer channel transistors are used to improve write performance. In addition, Ultra-short-segmented write in this structure improves write speed and robustness. The proposed architecture is pseudo-write-tolerable for half-selected bits, which allows bit interleaving. The half-selected bits during a write operation, as shown in Fig. 3.2(d), experience the same situation as a read, i.e. all pull down transistors are off and the access transistors are on. In addition read

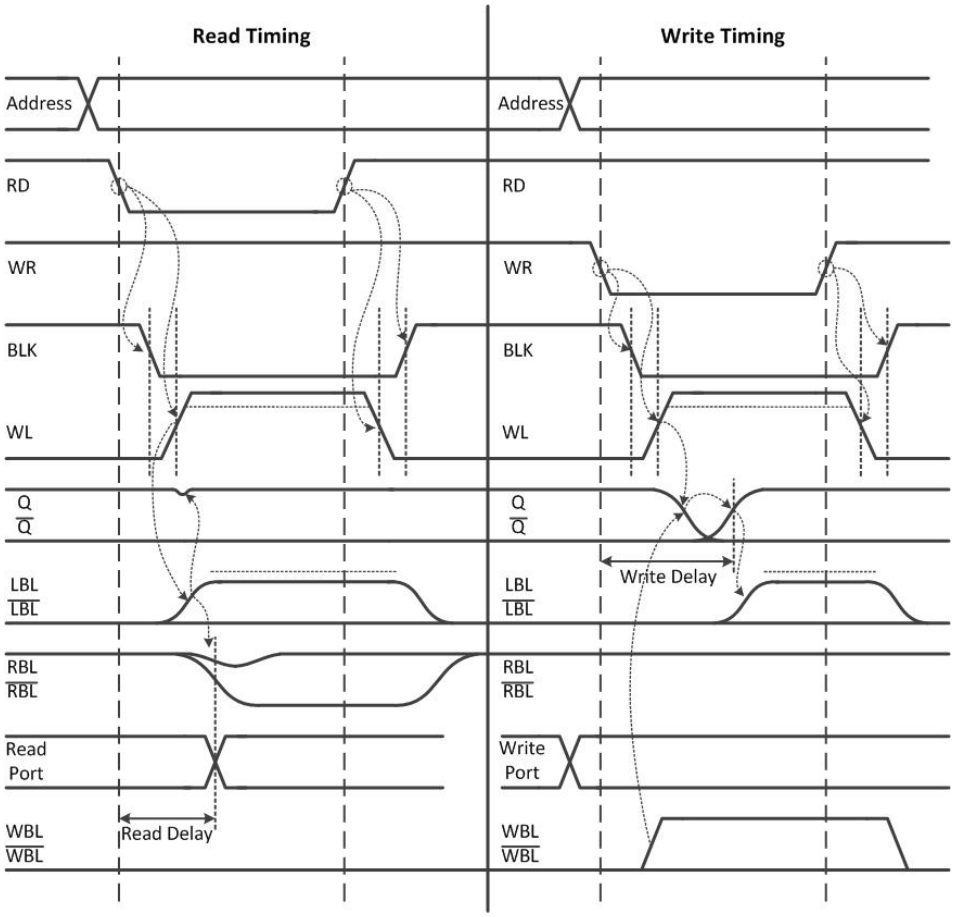


Fig. 3.5. (a) Read (b) Write operation timing diagram.

stability is carefully analyzed considering the effect of other bits on the same bitline, therefore a small, acceptable disturbance occurs on the half-selected bits during a write. The difference between half-select disturbance in the conventional 6T and the proposed architecture is the very short and low capacitance bitline in the proposed block which causes acceptable disturbance. As a result the proposed architecture allows bit-interleaving among columns and improves memory density.

The write operation in the proposed architecture is single-ended, i.e. no pull up available at the other side; however because of the distributed nature of the write driver, the write pull down transistors (Mwr_1 and Mwr_2) can easily be designed to overcome the pull up transistor (PMOS) in the selected bit and

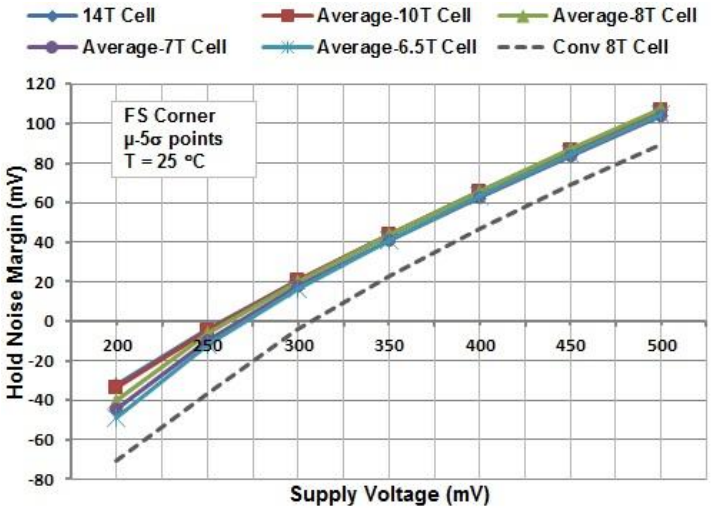
leakage from few unselected bits. In addition boosted wordline voltage will improve write strength. To verify this, we conduct a write noise margin analysis by sweeping WL voltage while considering the effect of other transistors. In this analysis, the whole block, i.e. including block mask, write, read transistors and other three unselected bits, initialized with the worst case data is used for simulation to consider all negative side effects during the simulation and have reliable results. Supply voltage of the block was set at 0.3 V and WL voltage was boosted by 50 mv, i.e. it was swept from 0 to 0.35 V. The Monte Carlo simulation results show a mean noise margin of 117.5 mV with standard deviation of 14.4 mV at room temperature and typical corner, which is far above zero and implies a reliable operation. The conventional two-sided write "without" boosting shows 151.5mV mean and 31.67mV standard deviation. Although the conventional method has higher mean, it has higher variations too. This analysis shows that proper write transistor sizing and wordline boosting will compensate for the single-ended write and allows a reliable write operation.

3.2.3 Leakage Reduction

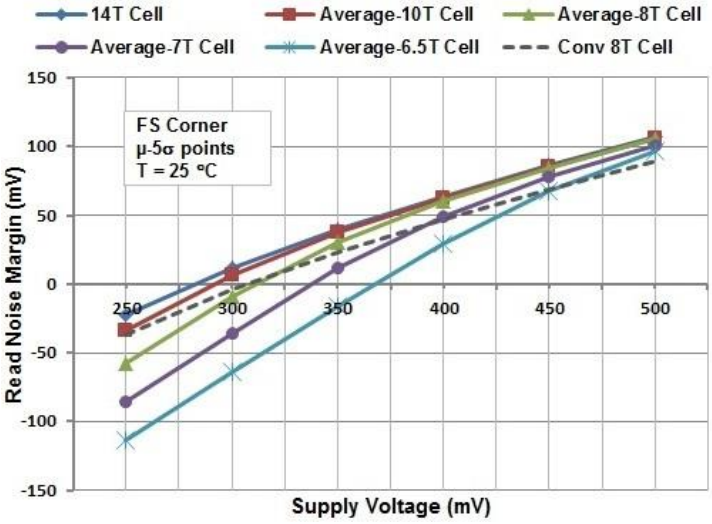
During a standby period, leakage current of the memory block can be further reduced by turning off Mpd_1 and Mpd_2 to put all local bitlines into a floating state, as shown in Fig. 3.2(e). In addition the global read bitlines are also set to a high impedance mode by turning off the pull up networks of the global bitlines. A chip select pin is used to control the standby mode of the whole memory macro.

3.3 Block Size Analysis

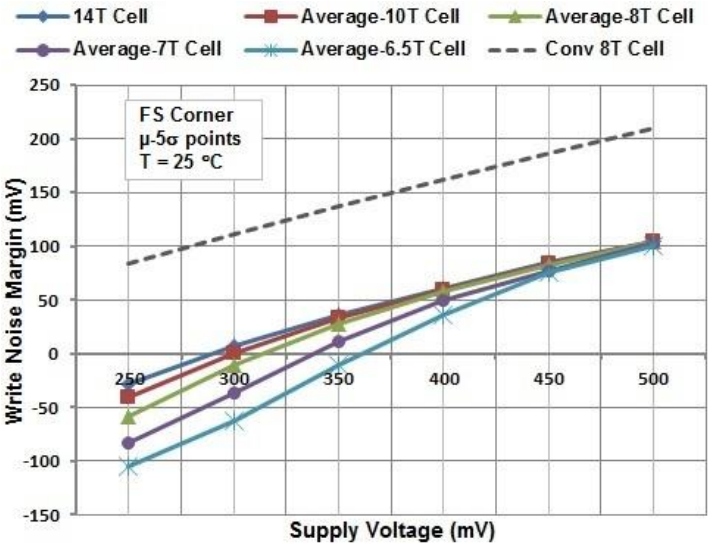
The proposed architecture is flexible in providing blocks of different number of bits, e.g. ranging from 1 to 16 bits per block as illustrated in Fig. 3.1(a) to (d). For various configurations, it is necessary to know the noise margins on hold, read, and write across different supply voltages and temperatures, as well as the area impacts of these configurations. These effects are investigated through simulations. In these simulations, first the process, i.e. global variations, is skewed by 3σ to obtain the worst case corner for noise margin which was Fast-Slow for this circuit. Then Monte Carlo simulation is performed at this process corner and μ-5σ point is considered as the edge of functionality. As stability of this architecture depends on the very short local bitlines, any analysis on hold, read, or write noise margins should consider side effects of few other bits and transistors connected to the local bitlines. As a result, for hold and read noise margins, the conventional method [56] was modified to calculate noise margin considering these side effects. In this analysis the whole block, same as the setup in Fig. 3.3, with the worst case initial data was used as load of the inverters. For write noise margin, the bitline sweep method is used to calculate noise margin [57]. In this simulation, only one local bitline was swept and the other side was left floating (driven only with data on the other side of the cell) to simulate the single-side write operation in this architecture. In addition, the whole block, same as Fig. 3.3, with the worst case initial data was used for this simulation. To make a comparison with the conventional 8T cell, the simulation results of a conventional 8T cell with transistor sizings same as [58] are added.



(a)



(b)

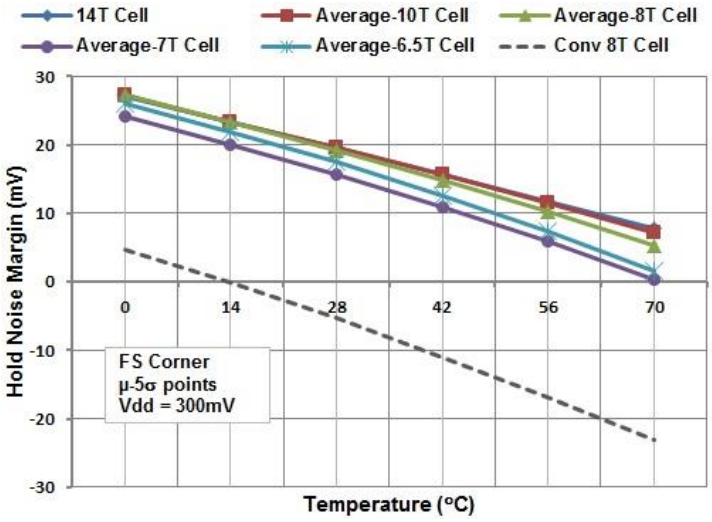


(c)

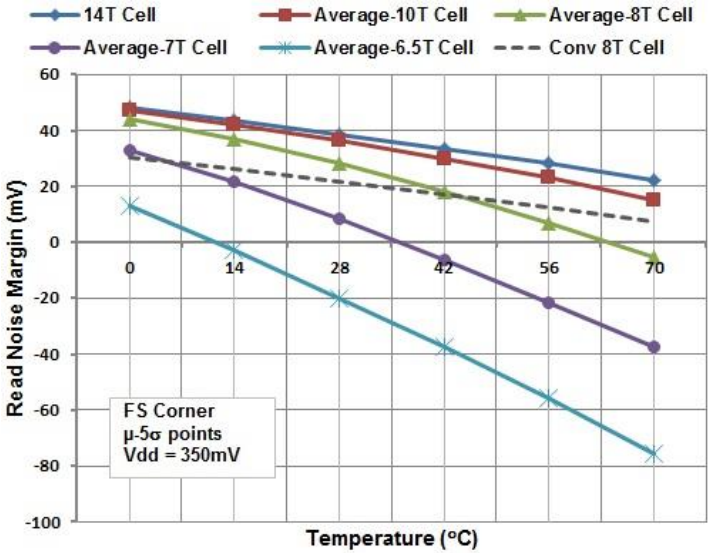
Fig. 3.6. Simulation results of (a) Hold (b) Read (c) Write noise margins versus supply voltage for different configurations of the proposed architecture and the conventional 8T cell.

We first investigate how the number of bits per block affects hold, read, and write noise margins at room temperature. The results are illustrated in Fig. 3.6(a), (b), and (c), respectively. In terms of hold noise margin, as can be seen from Fig. 3.6(a), increasing the number of bits has small impact on the stability of the block and it only increases minimum hold voltage by 13 mV from the 14T block to the average-6.5T block. Low dependency of hold noise margin on the number of bits is due to two series transistors connected between every two bits which reduce the impact of the storage nodes on each other. It should be noted that when the chip select is low, the block mask transistors, Mpd_1 and Mpd_2, are off. Despite hold noise margin, the number of bits per block has a considerable impact on read and write stability of the proposed architecture as can be seen from Fig. 3.6(b) and (c). This phenomenon happens because during a read or write operation, the access transistors of the intended data bit are on, therefore only one transistor separates the storage nodes in one block. As a result leakage of the access transistors increases substantially and causes higher dependency of read and write noise margins to the number of bits in the block with respect to the hold state.

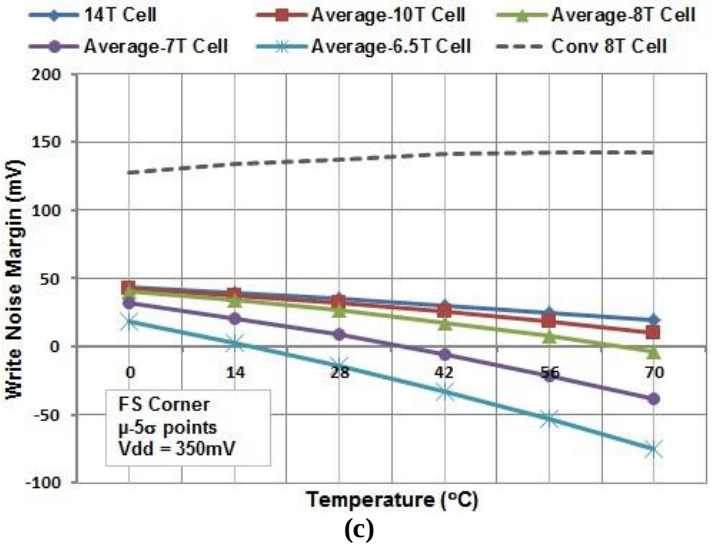
Varying the number of bits per block from 1 to 16, i.e. the 14T block to the average-6.5T block, increases minimum read and write operating voltage by 88 mV and 81 mV, respectively. These results clearly show why the standard 6T cell, which is an extension of this method with very large number of bits per block, fails at very low voltages. Comparing to the conventional 8T, the proposed cell has higher hold stability in order to be able to tolerate a small disturbance during a read. As a result, read stability is almost the same for both cases. Write stability is much stronger in the conventional 8T and that



(a)



(b)



(c)

Fig. 3.7. Simulation results of (a) Hold (b) Read (c) Write noise margins versus temperature for different configurations of the proposed architecture and the conventional 8T cell.

is the reason the conventional 8T has the pseudo-write problem in the half-selected cells. As we expect, the behavior of hold, read, and write noise margins are the same as the 14-T cell because they have fewer dependencies on the other cells.

Next, we study the effect of temperature on stability of various configurations. Simulations were performed at 300 mV supply voltage for hold noise margin and 350 mV for read and write noise margins. The results are shown in Fig. 3.7(a)-(c), respectively. Three interesting behaviors can be observed in these graphs. First, stability of the cell reduces by increasing temperature. This stability reduction is mainly due to the inverter gain and drain current reduction at higher temperatures in this technology. Second, similar to the number of bits per block dependency, read and write noise margins are more sensitive to temperature variations than hold noise margin, i.e. the graphs in Fig. 3.7(b) and (c) have a steeper slope than that of Fig. 3.7(a). During read and write operations, the access transistors are on and the storage node is connected to the local bitline and other access transistors. Besides, leakage current of other access transistors, i.e. write and block mask transistors, goes up with increasing temperature which introduces additional load to the storage node and further reduces stability of the cell with respect to the hold state. Third, increasing the number of bits per block increases dependency of hold, read, and write noise margins on temperature. This can be seen in Fig. 3.7 in terms of steeper slope from the 14T to the average-6.5T block in all graphs. This effect is more visible in read and write noise margins as they have only one off transistor between the storage nodes. By increasing the number of bits per block, the number of leaking elements, i.e. the access transistors of unselected

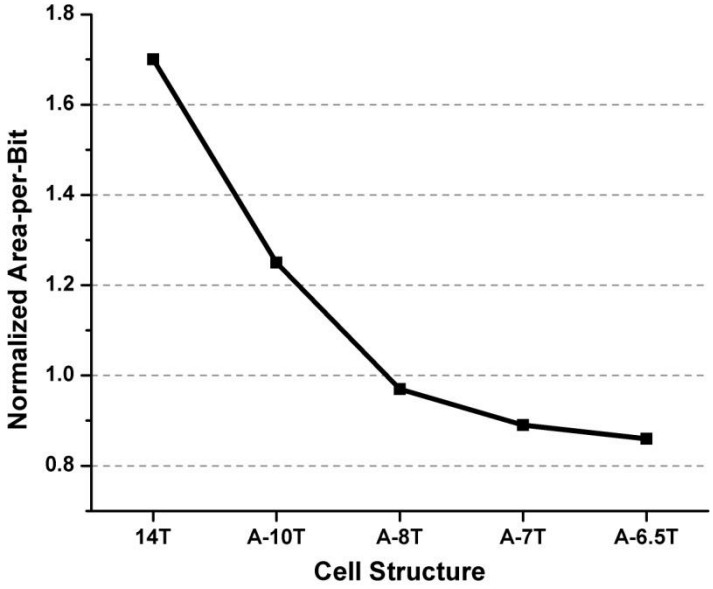


Fig. 3.8. Average area-per-bit of different configurations of the proposed architecture, normalized to the conventional 8T cell area reported in [32].

bits in the same block, increases. Positive temperature dependency of these leakages decreases cell stability faster in the blocks with larger number of bits per cell.

The last important point is the area efficiency of the block, i.e. average required area to store one bit. As we expect, despite noise margin, area efficiency improved with increasing the number of bits per block. Fig. 3.8 depicts normalized area versus the block configuration. Area in this graph is obtained by generating the layouts in the 0.13 µm CMOS process for each configuration and normalizing them to the area of a conventional 8T cell in the same technology [32]. As can be seen from Fig. 3.8, the area benefit saturates at the average-8T block, therefore this configuration can be an acceptable choice for most of ultra-low-voltage applications.

In summary, the selection of a block from various configurations is a trade-off among several factors, e.g. minimum operating voltage, area constraint, or minimum data retention voltage (DRV). Based on this analysis, to

reduce the minimum operating voltage, the number of bits per block should be kept as low as possible, i.e. the 14T or the average-10T block should be used. On the other hand, to have minimum area 8, 16 or 32 bits per block should be used. In terms of minimum DRV, as long as the number of bits per block is low, e.g. less than 16, the choice of configuration has a negligible effect on the minimum DRV.

3.4 Subthreshold Device Sizing Based on the Reverse Narrow-Width Effect (RNWE)

Shallow trench isolation (STI) was proposed in 1981 to improve packing density and electrical isolation of devices. Unlike local oxidation of silicon (LOCOS) isolation, this technique provides a sharp edge instead of a bird's beak shape. Due to the higher electric field at the edge of the channel, caused by fringing of the electric field of the gate as shown in Fig. 3.9, the threshold voltage is reduced by decreasing the channel width [59]. In other words, less electric field is required to form the channel. This phenomenon is called reverse narrow-width (or channel) effect (RNWE) which is more visible in NMOS transistors. The RNWE is observed in small geometries where the corner area affected by fringing electric field of the gate is comparable to the total width of the device.

In general this effect is undesirable as it increases the leakage current of circuits. However, on the further investigation, it reveals the possibility of utilizing the RNWE as a design parameter to optimize a subthreshold SRAM. Fig. 3.10 compares variation of the threshold voltage of a minimum-length NMOS transistor with respect to the transistor width at the super-threshold

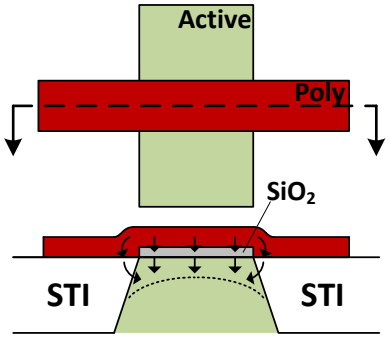


Fig. 3.9. Cross section of a MOSFET showing electric field in a narrow channel.

supply, i.e. 0.8 V, and the subthreshold supply, i.e. 0.35 V, in 65 nm technology. As can be seen from this figure, in both cases RNWE causes reduction in the threshold voltage. However, the drain current behaves completely different in these two regimes as shown in Fig. 3.11.

This behavior can be explained by the exponential dependency of the drain current on the threshold voltage in the subthreshold regime. By decreasing the width of a transistor, W/L ratio tends to decrease the current linearly. On the other hand, decrease in the threshold voltage tends to increase the current with quadratic ratio in the super-threshold and exponential ratio in the subthreshold. The final result depends on the net consequence of these two effects. As the change in the threshold voltage due to decreasing width is small and saturates early, the effect of decreasing W/L ratio dominates in the super-threshold regime and the drain current continuously decreases with decreasing width. In contrast, the small change in the threshold voltage translates to an order of magnitude change in the drain current due to the exponential dependency of the drain current on the threshold voltage in the subthreshold regime. Therefore the effect of the threshold voltage overcomes the effect of W/L ratio at narrow width. By increasing width, the increase in the threshold voltage starts to saturate, as shown in Fig. 3.10, as a result the effect of W/L ratio

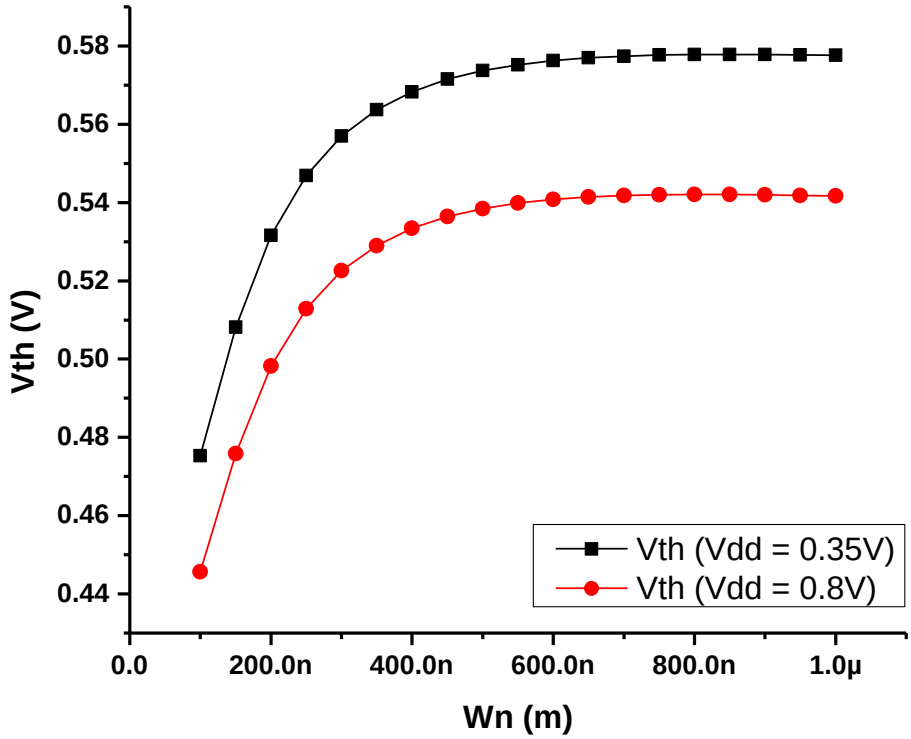


Fig. 3.10. Effect of transistor width on the threshold voltage of a minimum-length NMOS transistor at super-threshold supply, i.e. 0.8 V, and subthreshold supply, i.e. 0.35 V, in the 65 nm technology.

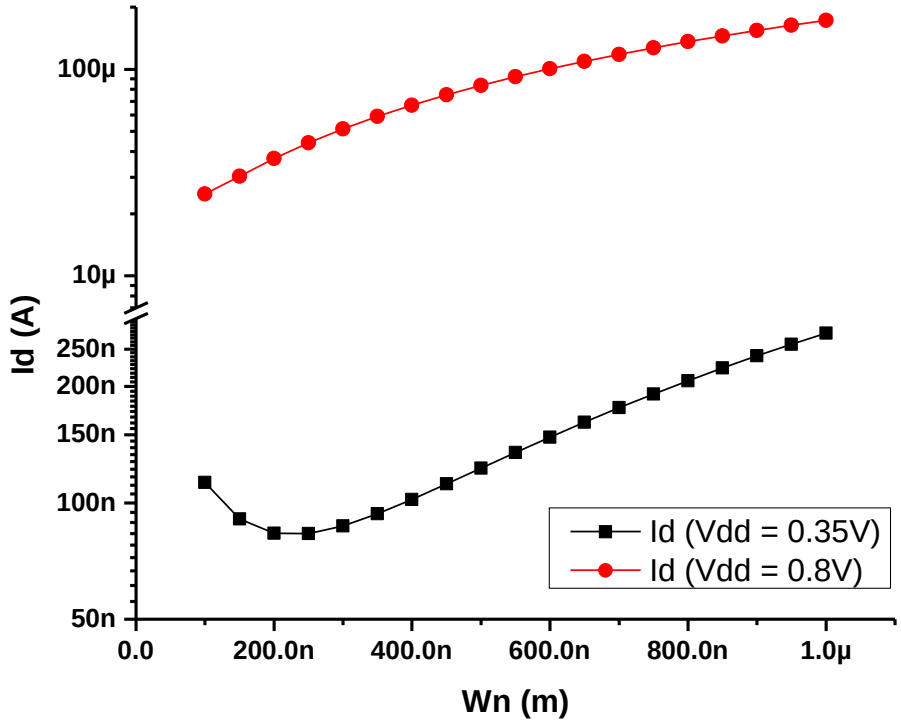


Fig. 3.11. Effect of transistor width on the drain current of a minimum-length NMOS transistor at super-threshold supply, i.e. 0.8 V, and subthreshold supply, i.e. 0.35 V, in the 65 nm technology.

outweighs the effect of the threshold voltage and the drain current starts to increase. Therefore, a minimum is observed in the drain current graph in the

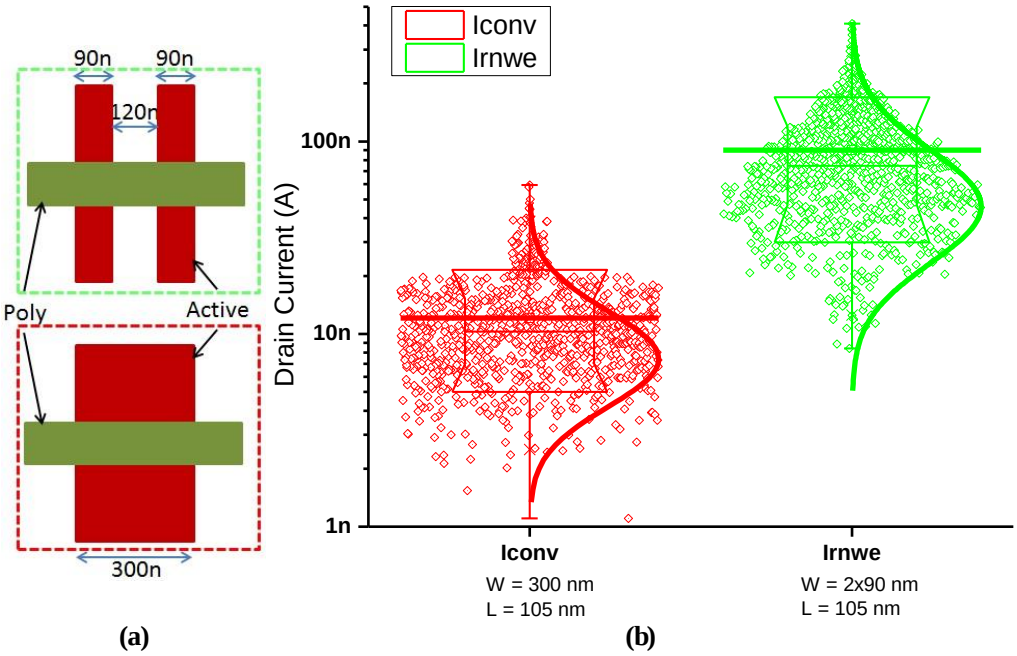


Fig. 3.12. (a) Iso-area transistors (b) Distribution of the drain current of the conventional sizing and RNCE-aware sizing under iso-area condition.

subthreshold regime as depicted in Fig. 3.11. This graph shows that in circuits like SRAMs in which most of the transistors are sized close to the minimum dimensions, a considerable change in width is required to make a transistor stronger in the subthreshold regime. Furthermore, it also suggests that a minimum-size transistor can be made stronger by increasing the number of fingers or multiplicity instead of directly increasing width. To evaluate this design technique, two transistors are compared under an iso-area condition as shown in Fig. 3.12(a) for a fair comparison. To observe the effect of variations, a Monte Carlo simulation was performed on each transistor and the drain current of two transistors were compared as shown in Fig. 3.12(b). As can be seen from this graph, although reducing width of the transistor increases the variation, RNWE-aware sizing provides 7.5× higher average drain current. Even most of the slow devices in RNWE-aware sizing have higher drain current than average of the conventional sizing. This analysis shows that increasing multiplicity of the transistor instead of increasing width can be an effective

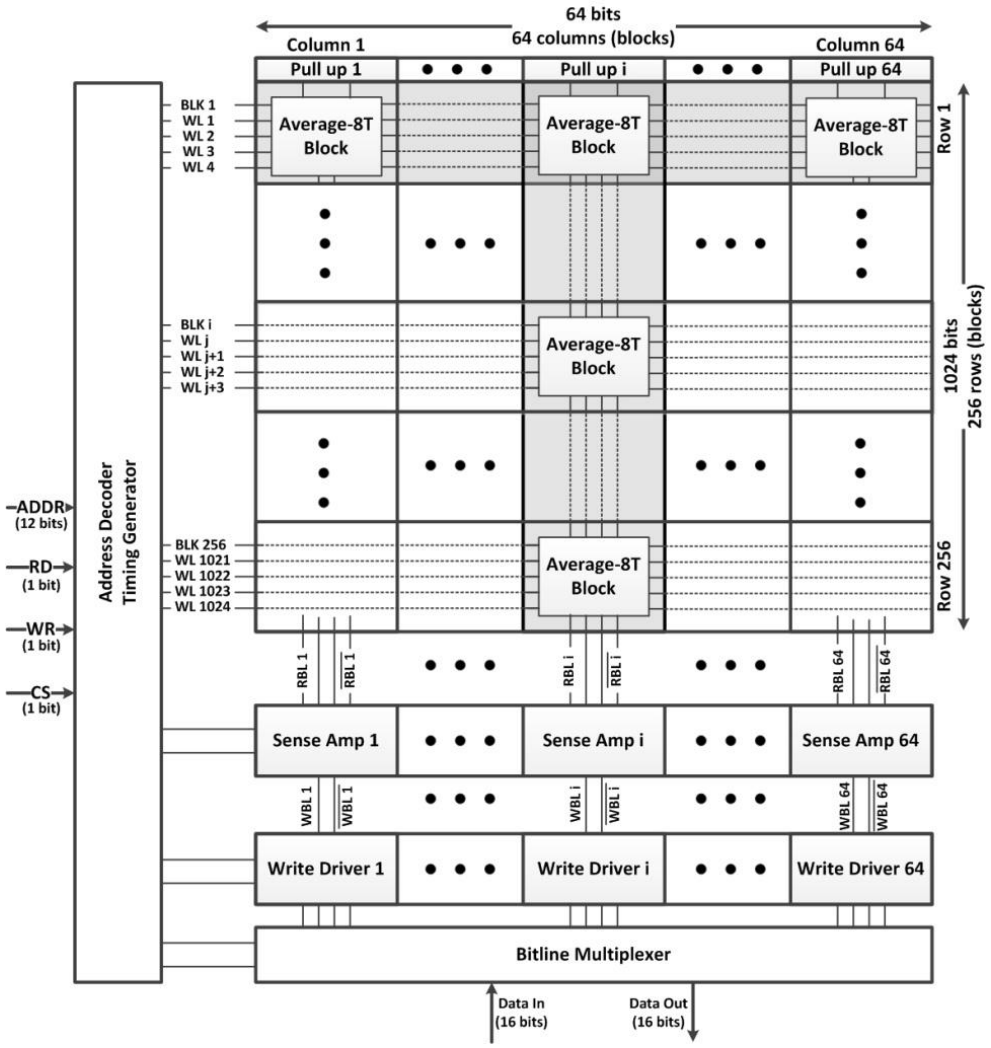
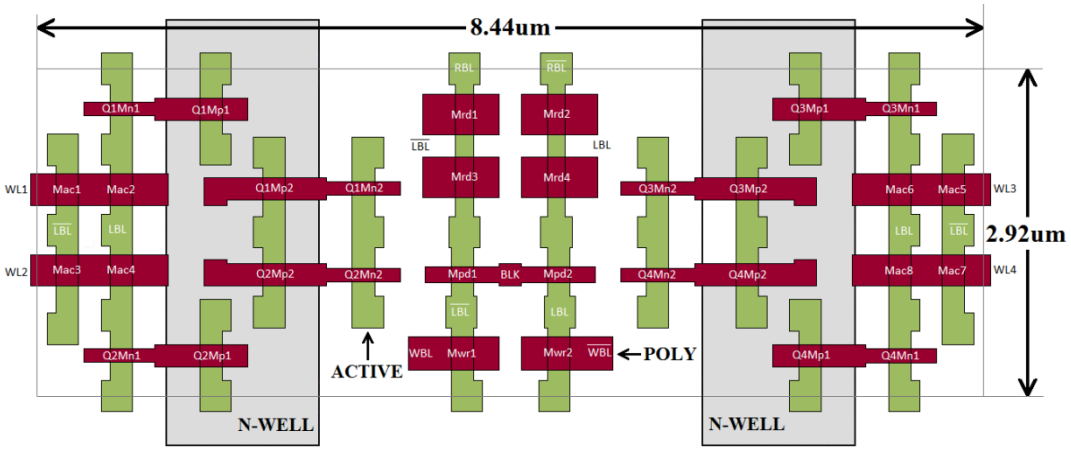


Fig. 3.13. Array-level block diagram of the 64 kb memory.

knob to make a transistor stronger in the subthreshold regime. In this work, RNWE-aware sizing is applied to the critical transistors in the write and read operations of the proposed average-8T architecture to improve the performance with acceptable area overhead. As a result, the access transistors were selected as they affect both read and write operations. According to simulations, the write operation requires further assistance; therefore the RNWE is also applied to the write transistors.



Name	Q1-4	Q1-4	Mac_{1-8}	$Mpd_{1,2}$	$Mwr_{1,2}$	Mrd_{1-4}
Type	NMOS	PMOS	NMOS	NMOS	NMOS	NMOS
Width	200n	200n	200n	200n	200n	160n
Length	120n	200n	280n	140n	300n	360n

Fig. 3.14. Layout and transistor sizing of the average-8T write/read decoupled block in the 0.13µm process.

3.5 Chip Implementation and Measurement Results

Three chips have been designed and fabricated in this work. First of all, to validate and show the effectiveness of the proposed architecture, two asynchronous SRAM memories with size of 16 and 64 kb, respectively, have been designed and fabricated in the 0.13 µm bulk CMOS process based on configuration of Fig. 3.1(a). Second, another 16 kb memory block utilizing reverse narrow-width effect was designed and fabricated in the 65 nm bulk CMOS process. The average-8T configuration was selected to balance between the minimum area and the minimum operating voltage. Fig. 3.13 shows the architecture of the implemented 64 kb memory block. This memory block has 64 columns and 1024 rows. A multiplexer selects 16 I/O bits out of 64 columns. The architecture of the 16 kb block is the same with the half number of columns, i.e. 32, and the half number of rows, i.e. 512. Fig. 3.14 shows the layout of the 64 kb block and sizing of its transistors. The average area per bit of this

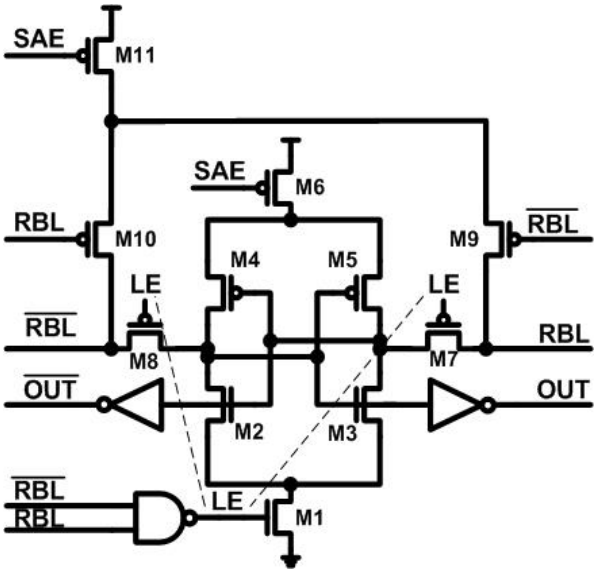


Fig. 3.15. The self-timed decoupled differential sense amplifier.

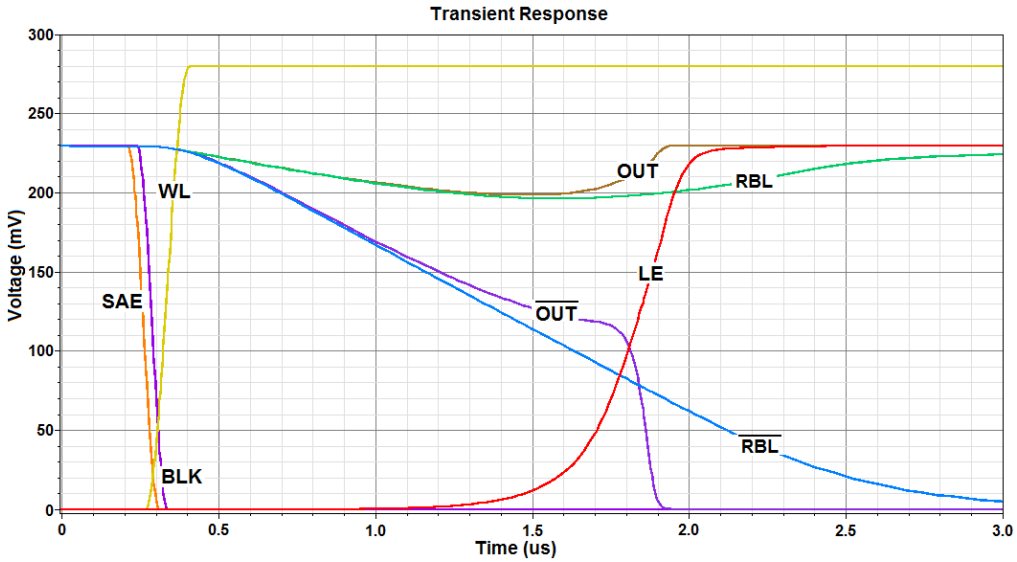


Fig. 3.16. Transient responses of the self-timed decoupled differential sense amplifier at 230mV and typical corner.

block is 3.2% smaller than a conventional 8T cell in the same technology [32]. In addition, this layout allows wider column pitch which is helpful in designing a better sense amplifier.

A self-timed differential sense amplifier [20] has been modified, as shown in Fig. 3.15, for this memory block to mitigate the effect of high variations in the subthreshold regime. Two decoupling transistors (M_7 and M_8) are added in the inputs of the sense amplifier to decouple the sense amplifier

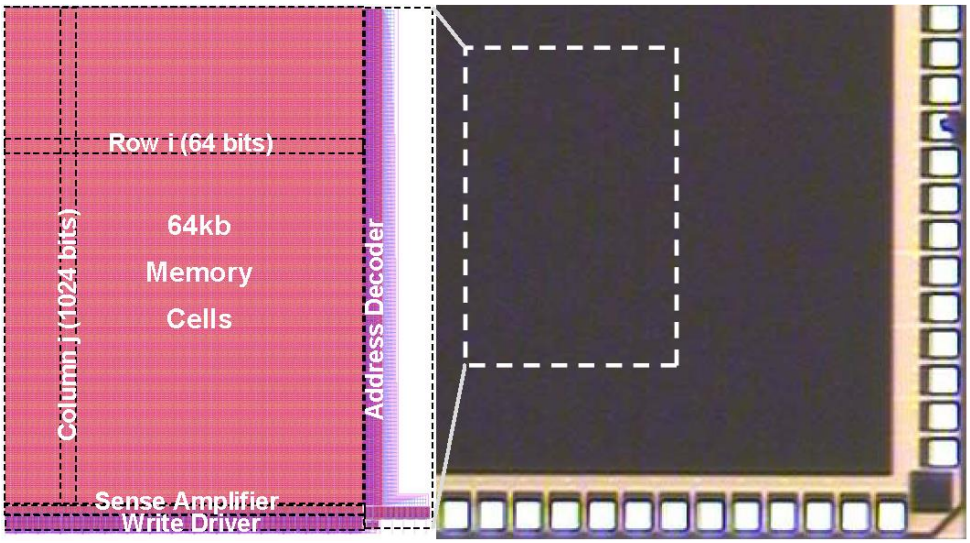


Fig. 3.17. (a) 64kb SRAM block layout. (b) The fabricated chip photo.

inputs from the high capacitance nodes (RBL and $\overline{\text{RBL}}$) in order to improve the read speed. Fig. 3.16 shows the transient response of the sense amplifier at 230 mV supply. In this circuit when sufficient voltage (more than sense amplifier offset) is developed across RBL and $\overline{\text{RBL}}$, the LE goes high via a skewed NAND gate. The NAND gate is designed to be triggered with voltages close to VDD. The LE pulse enables the sense amplifier/latch and at the same time disconnects the sense amplifier inputs from the high capacitance nodes for a faster latching. A statistical analysis has been performed to make sure that the developed voltage across the sense amplifier input is more than sense amplifier offset.

Fig. 3.17 shows the layout and die photograph of the 64 kb chip. The total chip area is 640 µm × 800 µm which is 16% less than that of the conventional 8-T RD SRAM with the same size and technology [32]. The area saving is due to smaller cell size and data-independent leakage characteristics of the proposed block which allow implementing a large number of bits per bitline, e.g. 1024 bits in this design, and removes the additional sense amp, write

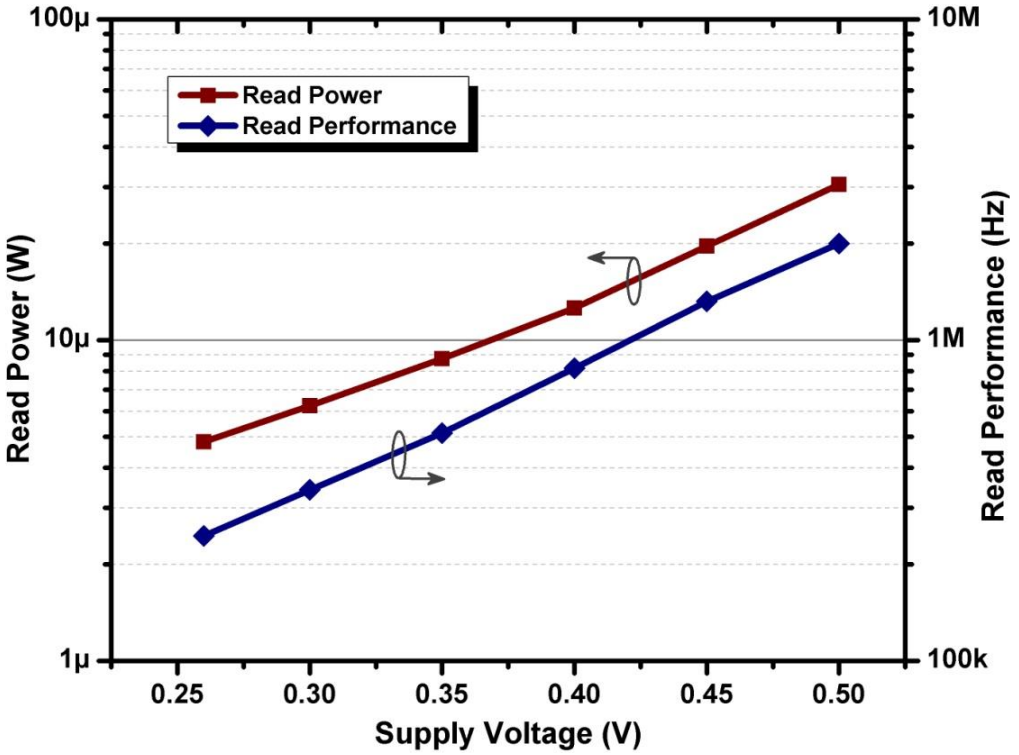


Fig. 3.18. Read power and performance versus supply voltage (T=25°C) for the 64 kb chip.

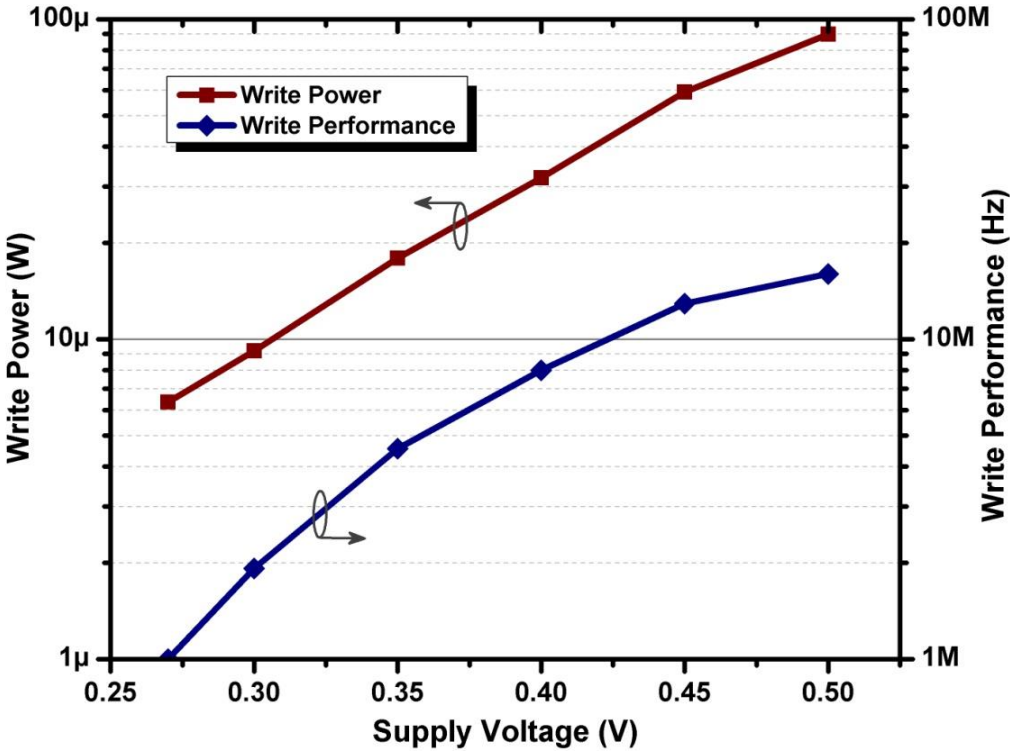


Fig. 3.19. Write power and performance versus supply voltage (T=25°C) for the 64 kb chip.

driver, and routings overhead. Note that the charge pump circuit is not included in this design. The rectangular area next to the address decoder at the right

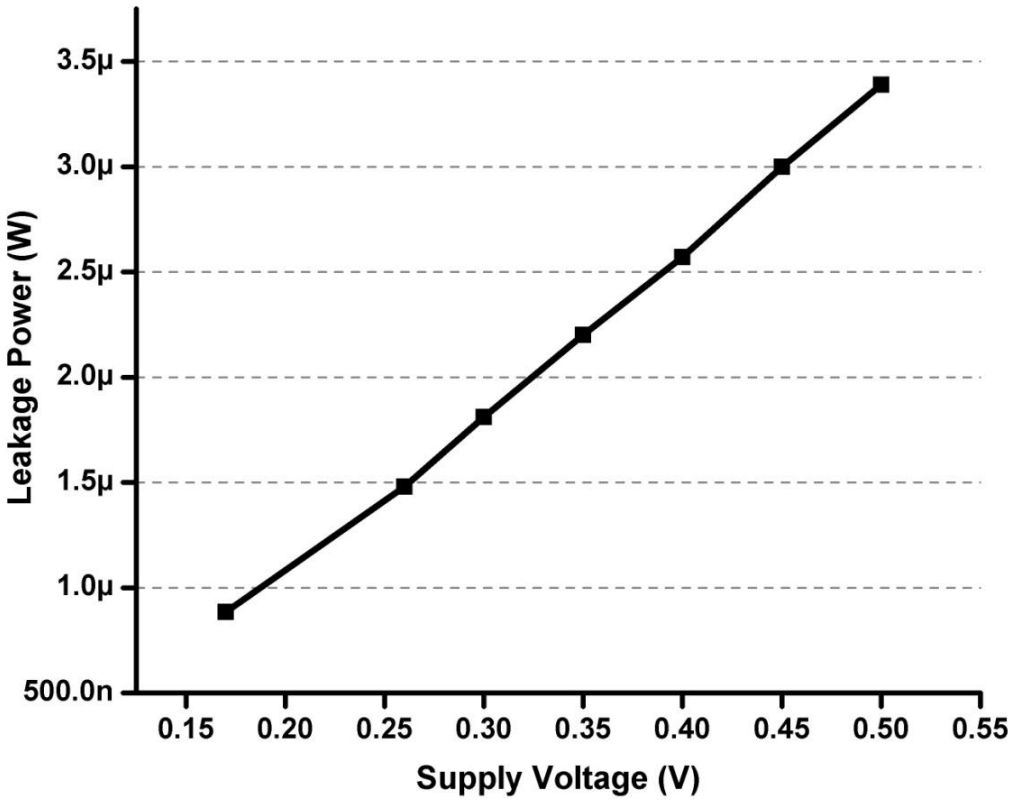


Fig. 3.20. Leakage power versus supply voltage at room temperature (T=25°C) for the 64 kb chip.

side should be sufficient for a charge pump circuit.

Fig. 3.18 and Fig. 3.19 show read and write speed and power at various supply voltages, respectively. The minimum read voltage is 260 mV at 245 kHz. The minimum write voltage is 270 mV at 1 MHz. Higher write speed is due to the distributed nature of the write drivers which facilitates this higher speed for the write. Fig. 3.20 shows leakage power variation with different supply voltages. The leakage current is measured with an asserted chip select which puts the whole memory in the standby mode. All measurements are performed at room temperature (25 °C). In the standby mode it can hold data down to 170 mV while consuming only 884 nW.

Energy consumption per read and write are depicted in Fig. 3.21. As can be seen, the minimum energy point occurs at the near-threshold which is

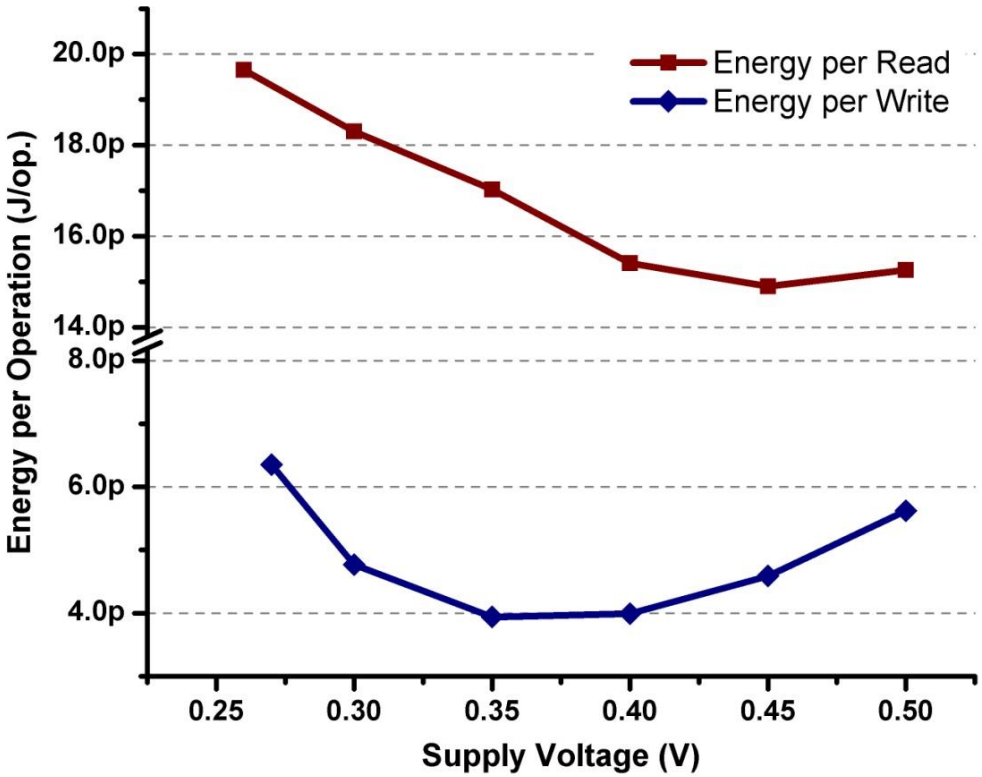


Fig. 3.21. Energy per read and write versus supply voltage (T=25°C) for the 64 kb chip.

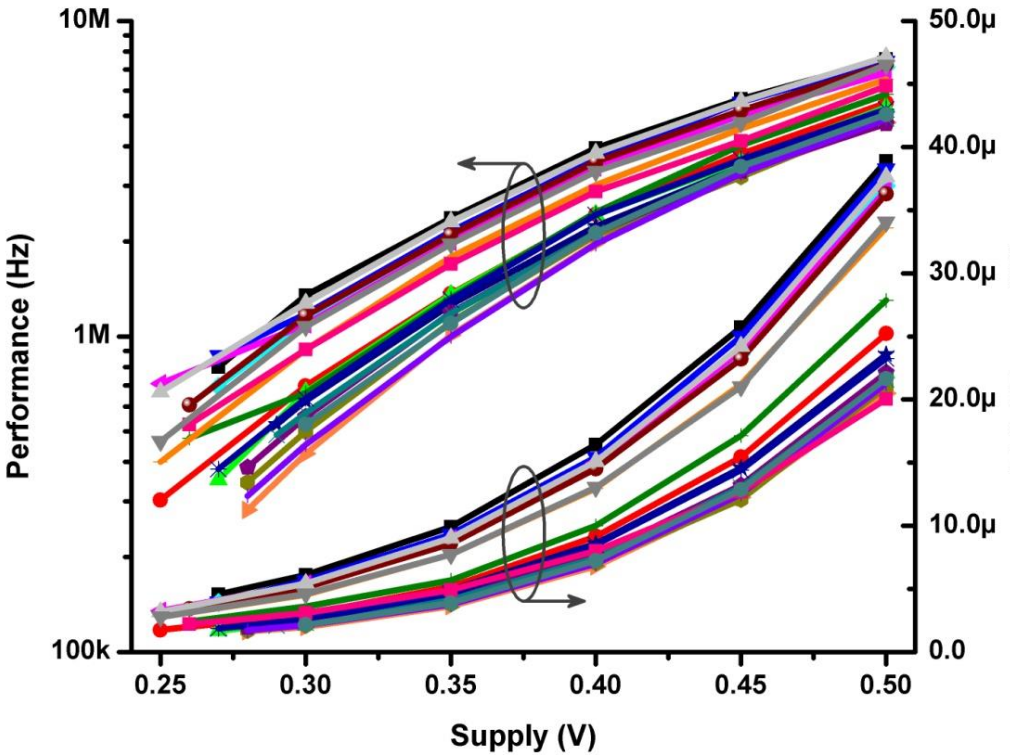


Fig. 3.22. Distribution of power consumption and performance versus supply voltage for the 16 kb chip, measured across 20 chips at room temperature.

14.9 pJ/read at 450 mV for a read and 3.9 pJ/write at 400 mV for a write. For read operations, 64 bits are read at one time leading to the minimum energy

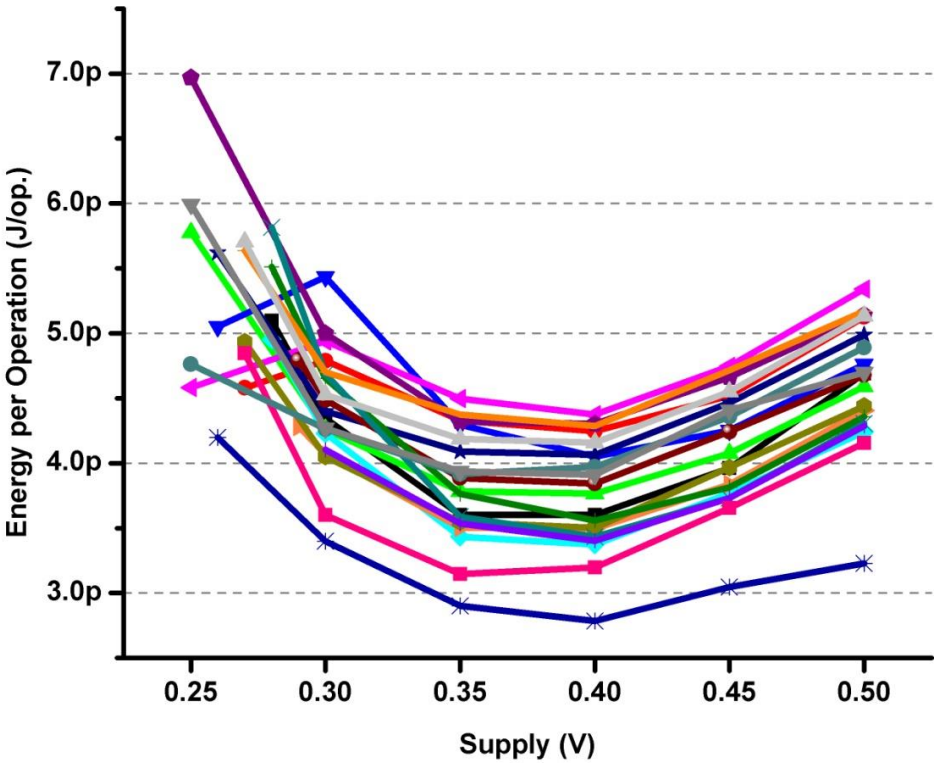


Fig. 3.23. Distribution of average energy-per-operation versus supply voltage for the 16kb chip, measured across 20 chips at room temperature.

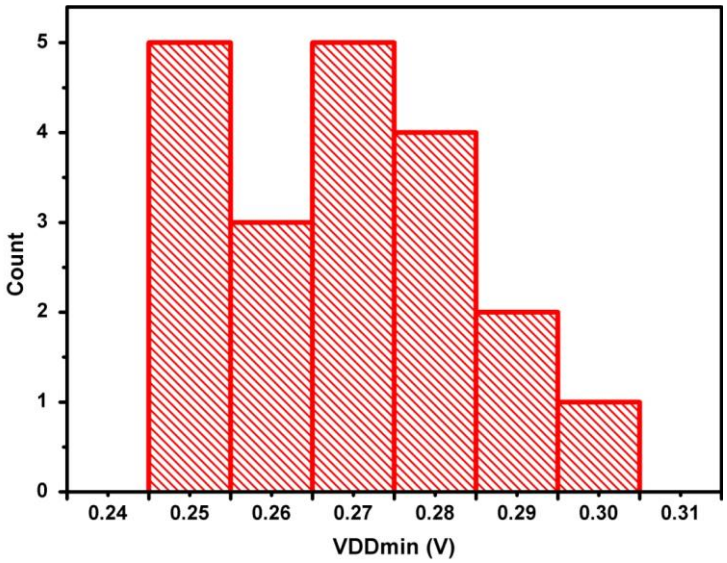


Fig. 3.24. Distribution of minimum operating voltage of 16 kb chip across 20 chips at room temperature.

per bit per read of 0.23 pJ. In the write operation 16 bits are written at one time, resulting in the minimum energy per bit per write of 0.24 pJ. It is possible to further reduce the energy per read if a small number of blocks are attached to each bitline, which speeds up the read operation. However, this approach increases area. Due to the limited number of dies for the 64 kb chip,

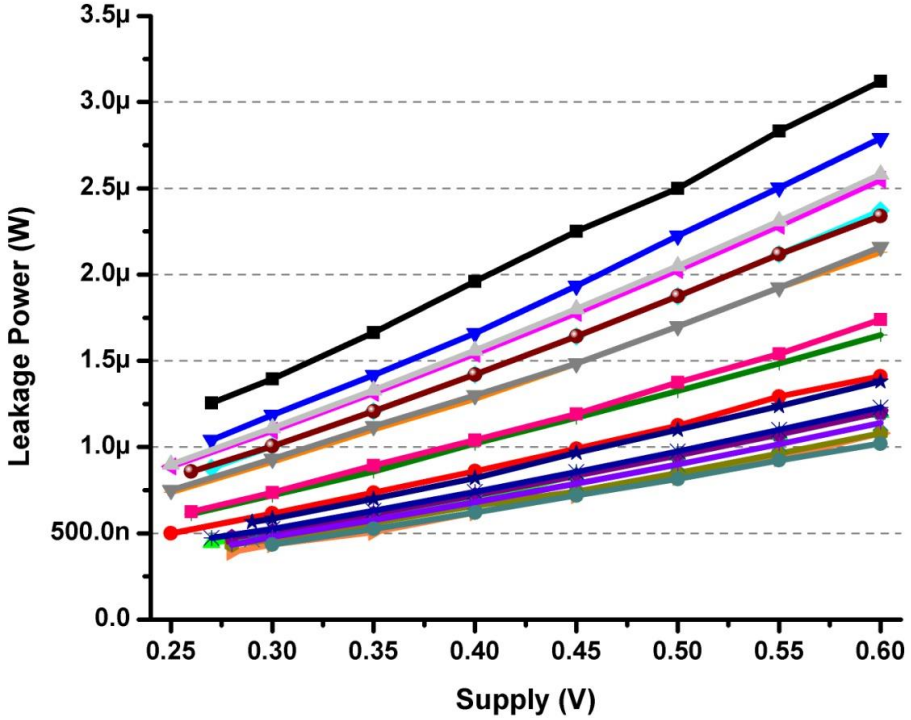


Fig. 3.25. Distribution of leakage power consumption versus supply voltage for the 16 kb chip, measured across 20 chips at room temperature.

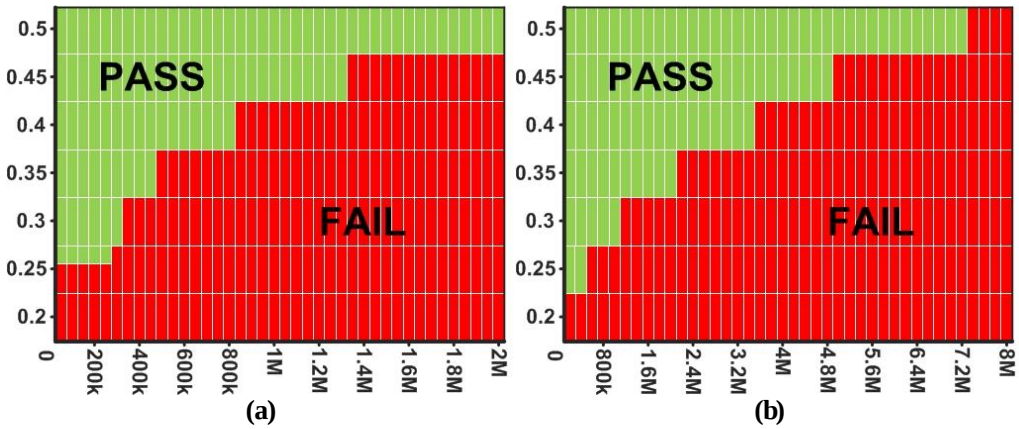


Fig. 3.26. Shmoo plot of (a) 64kb (b) 16kb SRAM blocks.

we designed and fabricated a 16kb chip in the same technology to further investigate the distribution of performance and power consumption. 20 pieces of 16kb chips were measured. Fig. 3.22 shows the distribution of read/write speed and total power consumption versus supply voltage.

As we expected, this block is faster than the 64 kb block and shows more efficient read and write. The minimum energy-per-operation point ranges from 2.8-4.4 pJ, as shown in Fig. 3.23. The distribution of the minimum

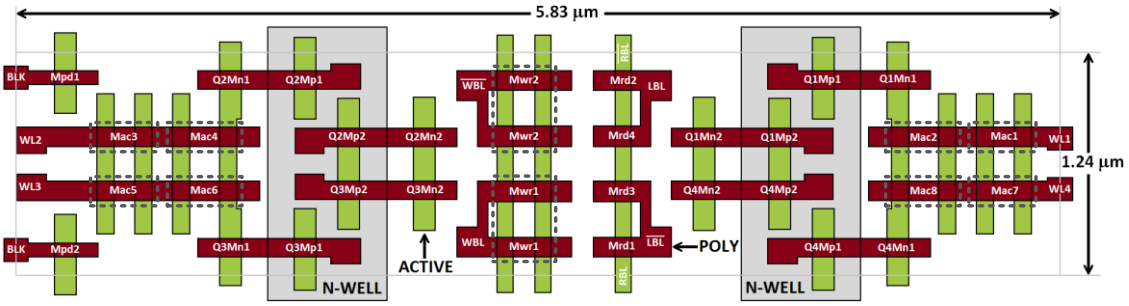
fully functional operating voltage is shown in Fig. 3.24. As can be seen from this graph, all 20 chips were functional down to 300 mV.

The leakage power distribution of these 20 chips is shown in Fig. 3.25. As expected in the subthreshold regime, the variation causes 2-3x change in leakage power (and in general for power and performance) of the memory block. However these results show a reliable operation of the designed block in the subthreshold regime. Fig. 3.26 depicts the Shmoo plot of the 64kb and 16kb blocks. Table 3.1 summarizes the specifications and performance of the fabricated chips and compares them with the recent subthreshold SRAM designs. This table shows that the proposed block has the smallest area and speed improvement with acceptable leakage power.

Table 3.1. Performance and Specifications Summary and Comparison.

	JSSC09 [32]	**TVLSI12 [24]**	**JSSC09 [12]**	**This work**	
Technology	0.13 µm	0.13 µm	65 nm	0.13 µm	
Memory Size	64 kb	2 kb	64 kb	64 kb	16 kb
Area	0.612 mm^2	1.063 mm^2	1.4* mm^2	0.512 mm^2	0.155 mm^2
Normalized Area	9.34 μm^2	519 μm^2	N.A.	7.81 μm^2	9.46 μm^2
Cell Structure	8T	Schmitt-Trigger	8T	average-8T	
No. of Bits per Bitline	512	NA	64	1024	512
Min. Supply	0.26 V	0.32 V Best: 0.15 V	0.25 V	0.26 V for read 0.27 V for write 0.17 V for hold	Average: 0.269 V Std: 15 mV Best: 0.25 V
Leakage Power	1.33 µW @ 0.23 V	0.11 µW @ 0.3 V	0.45 µW @ 0.25 V	0.88 µW @ 0.17 V	Average: 0.756 µW @ 0.3 V Best: 0.508 µW @ 0.3 V
Performance	100 kHz @ 0.23 V 15 MHz @ 0.6 V	270 kHz @ 0.3 V	20 kHz @ 0.25 V 50 kHz @ 0.3 V 500 kHz @ 0.4 V	Read: 245 kHz @ 0.26 V 2 MHz @ 0.5 V Write: 1 MHz @ 0.27 V 16 MHz @ 0.5 V	Average: 820 kHz @ 0.3 V 6 MHz @ 0.5 V Best: 2.37 MHz @ 0.3 V 8 MHz @ 0.5 V

* Die size



Name	Q1-4	Q1-4	Mac_{1-8}	$Mpd_{1,2}$	$Mwr_{1,2}$	Mrd_{1-4}
Type	NMOS	PMOS	NMOS	NMOS	NMOS	NMOS
Width	120 n	120 n	2×90 n	120 n	4×90 n	90 n
Length	105 n	105 n	105 n	80 n	110 n	110 n

Fig. 3.27. The layout and transistor sizing of the RNWE-aware average-8T block in the 65 nm process.

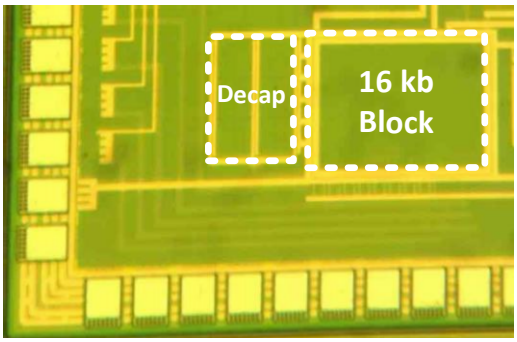


Fig. 3.28. Die photograph of the fabricated 16 kb chip in the 65 nm technology.

Another 16 kb memory block in the 65 nm CMOS technology utilizing the reverse narrow-width effect was designed and fabricated. This block has 32 columns and 512 rows. A multiplexer selects 8 bits out of 32 bits to limit the number of pads in the chip. Fig. 3.27 shows the layout and transistor size of the bit-cell block used in this memory. Gate length is chosen to be longer than the minimum size to reduce leakage current and variations. This block stores 4 bits, therefore the average area per bit of this bit cell is 1.81 μm^2. As can be seen from the layout and sizing, to make the write and access transistors stronger, multiple copies of a narrow-width transistor were used instead of just increasing width of the transistor. These transistors are marked with a dashed-box in the layout for clarification. Fig. 3.28 shows the die photograph of the implemented chip. The chip area is 190 µm × 250 µm therefore the

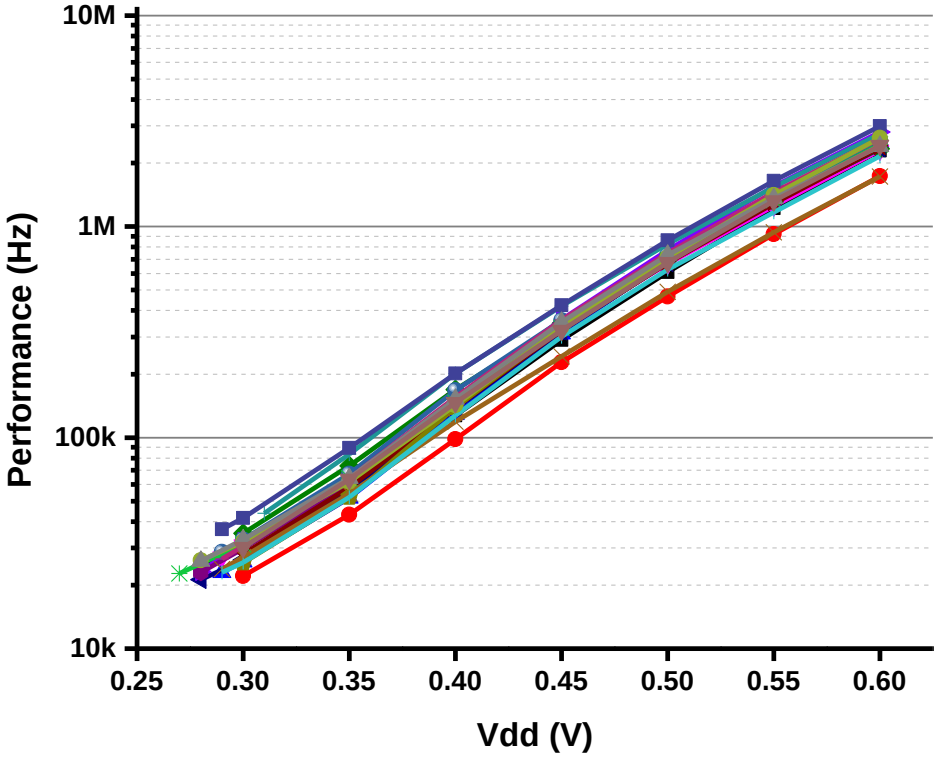


Fig. 3.29. Performance at various supply voltages measured for 20 chips.

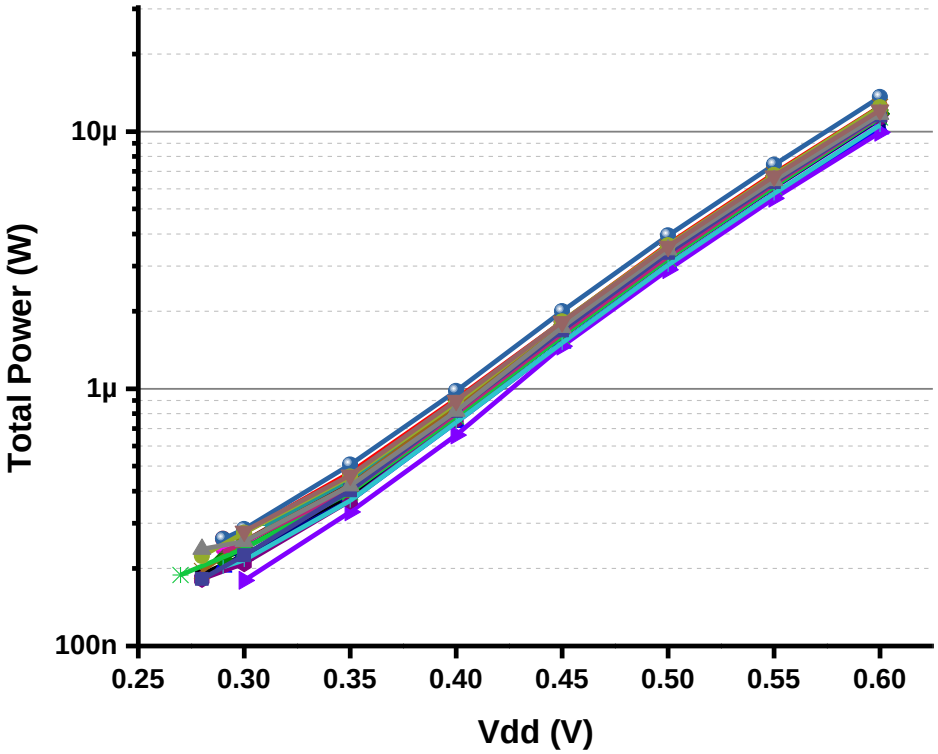


Fig. 3.30. Total power consumption at various supply voltages measured for 20 chips.

average area per bit considering all peripheral overheads is 2.97 μm^2 which is smaller than the other 65 nm subthreshold designs as compared in Table 3.2.

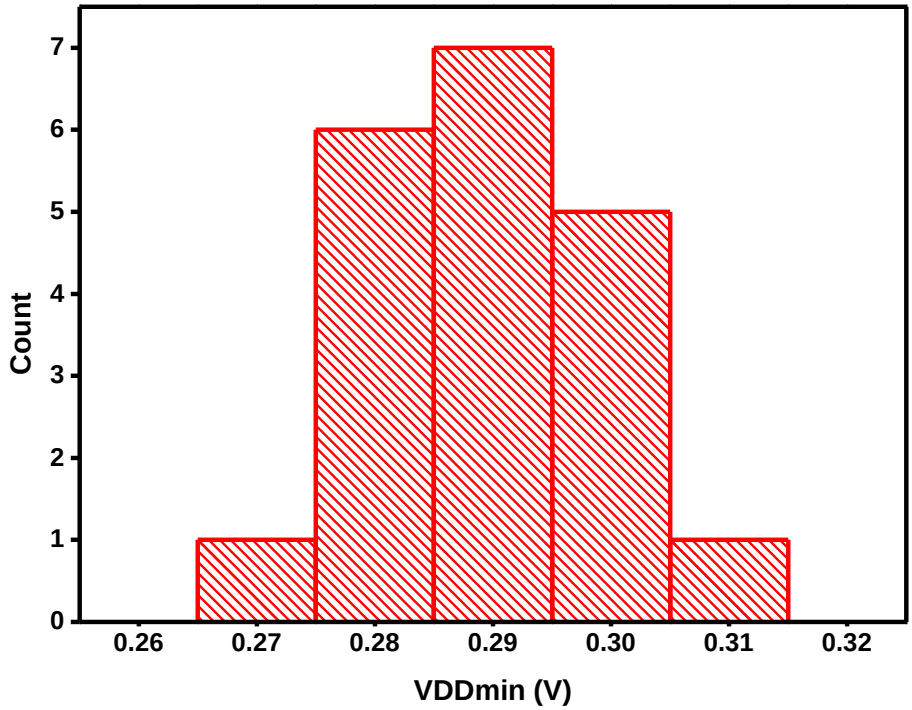


Fig. 3.31. Distribution of the minimum fully functional supply voltage, measured for 20 chips.

The maximum operating speed and total power consumption of the chip at this speed at various supply voltages were measured for 20 chips and depicted in Fig. 3.29 and Fig. 3.30, respectively. As can be seen from these graphs, on average, the chip is working from 30.8 kHz at 0.3 V while consuming 246 nW up to 2.42 MHz at 0.6 V while consuming 11.6 µW.

All 20 chips are functional down to 0.31 V as shown in the minimum functional supply voltage distribution in Fig. 3.31. Leakage power of the chip in the standby mode at various supply voltage and its distribution across 20 chips is depicted in Fig. 3.32. On average this chip consumes only 105 nW at 0.3V in the standby mode. Comparing this value with the total power consumption at the full speed reveals that half of the total power is consumed by the leakage of the cell which shows the importance of leakage power management. In the best case, this chip consumes only 64.4 nW at 0.28 V in the standby state.

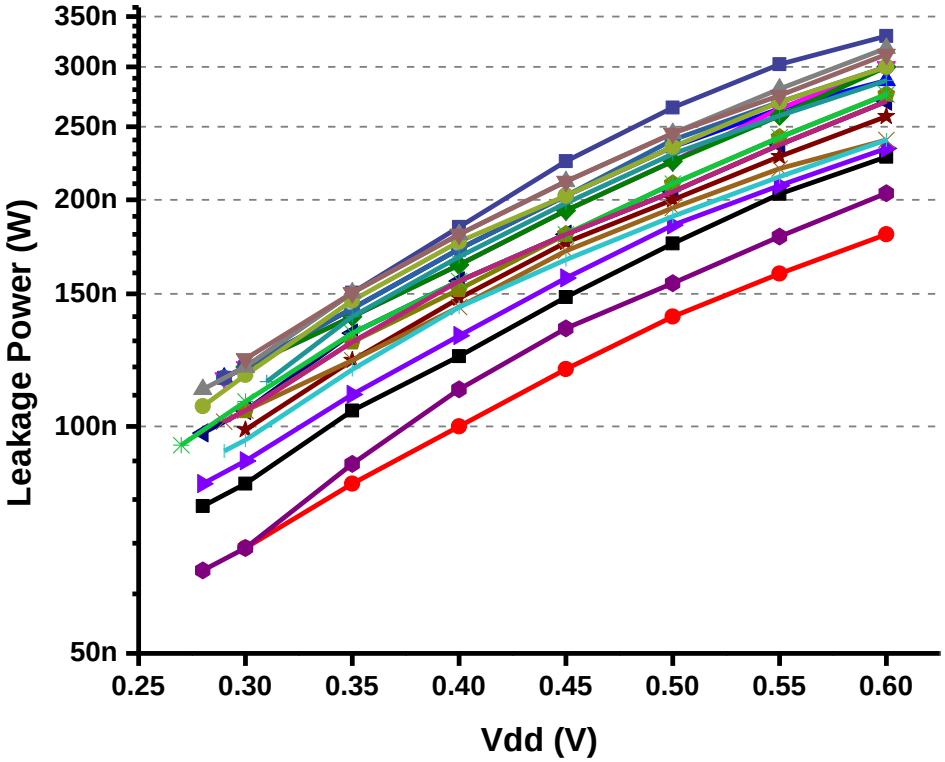


Fig. 3.32. Leakage power at various supply voltages measured for 20 chips.

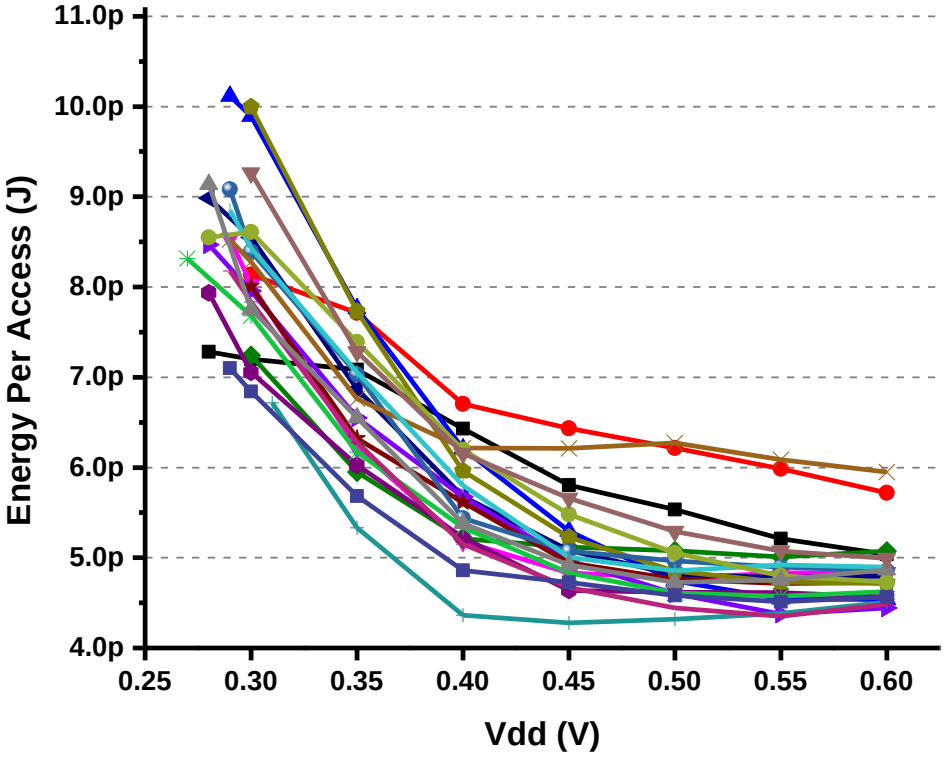


Fig. 3.33. Minimum energy per access at various supply voltages, measured for 20 chips.

The distribution of the minimum energy per access is shown in Fig. 3.33. The minimum energy per access occurs at 0.55 V to 0.6V in which the

chip consumes around 4.85 pJ. In the best case, the chip consumes only 4.28 pJ/access.

Table 3.2 summarizes the specifications and performance of the chip and compares it with the recent subthreshold SRAM designs in the same technology. As can be seen from this table, this work achieves the lowest leakage power per bit, and the lowest minimum energy per access compared with the other subthreshold SRAMs in the same technology. Furthermore, thanks to the area-efficient average-8T architecture, the lowest average area per bit, considering all peripheral overheads, is also reported in this work.

Table 3.2. Performance and Specifications Summary and Comparison of 65nm chip.

	JSS08 [11]	JSSC09 [12]	JSSC12 [30]	ISLPED12 [60]++	This work+
Technology	65 nm	65 nm	65 nm	65 nm	65 nm
Block Size	256 kb	64 kb	72 kb	1 kb	16 kb
Total Area	1.89×1.12 mm^2*	1.4×1 mm^2*	560×400 μm^2	182×45 μm^2	255×191 μm^2
Cell Structure	8T	8T	9T	9T	Average-8T
No. of Bits per Bitline	256	64	64	16	512
Min. Supply	0.35 V	0.25 V	0.35 V	0.3 V	All: 0.31 V Best: 0.27
Maximum Speed	25 kHz @ 0.35 V	50 kHz @ 0.3 V	229 kHz @ 0.35 V	38.46 kHz @ 0.3 V	Average: 62 kHz Best: 89 kHz @ 0.35 V
Leakage Power	1.65 μW @ 0.3 V	0.45 μW @ 0.25 V	2.29 μW @ 0.275 V	105.9 nW @ 0.3 V	Average: 105 nW Best: 69 nW @ 0.3 V
Leakage Power per Bit	6.29 pW @ 0.3 V	6.87 pW @ 0.25 V	31.06 pW @ 0.275 V	103 pW @ 0.3 V	Average: 6.41 pW Best: 4.21 pW @ 0.3 V
Area per Bit with Overheads	N.A.	5.7 μm^2**	3.038 μm^2	8.1 μm^2	2.97 μm^2
Min. Energy per Access	N.A.	11 pJ	4.5 pJ	5.82 pJ	Average: 4.85 pJ Best: 4.28 pJ

+ Average values are average of 20 chips.
++ Simulation results
* Die area
** Estimated from chip photo

3.6 Summary

In this chapter, a new average-8T write/read decoupled (A8T-WRD) SRAM cell architecture for the sub/near-threshold embedded SRAM has been proposed and its various configurations were analyzed. Measurement results for two prototype chips in 16 and 64kb show that the proposed block allows a faster and more robust read and write without any trimming or assistant circuits (except 50 mV WL voltage boost) while it reduces the total area by 16%. The performance gain is due to the introduction of the decoupled write port, data independent leakage, differential read port, and grouped data bits. In addition the proposed technique allows bit interleaving in the wordline which yields a denser memory. Furthermore, utilizing reverse narrow-width effect in sizing of the subthreshold memory has been proposed. This technique has been verified by implementing another 16 kb memory block in the 65 nm CMOS technology.

Chapter 4
Low-Power Microcontroller for Biomedical Applications

4.1 Overview

The development of the ECG SoC has attracted much attention in the recent years [41, 61-63]. Some of the ECG SoCs [41, 62, 63] integrate microcontroller on the chip to conduct pre-processing in order to minimize the wireless transmission of raw data. Commercial low-power microcontrollers, e.g. PIC and MSP430, and ultra-low-power implementations in [41, 46] are not the power-efficient choices for the target application as they need abundant clock cycles to implement a common task like filtering. On the other hand, more powerful DSPs [53] are too power hungry and thus not suitable for a power-limited biomedical SoC. As a result, a low-power microcontroller core with a few carefully-selected DSP features is a more appropriate choice for wireless biomedical sensors. In addition, biomedical sensors often demand a large block of memory to facilitate a burst mode wireless transmission. In

some of the previous works [41, 64], the memory is placed in a separate and higher voltage domain, which increases power consumption and requires additional supply. In this chapter, we present an ultra-low-power microcontroller and an ECG SoC based on this microcontroller that address issues in the existing solutions. This chapter is organized as follows. In Section 4.2, the architecture of the proposed microcontroller is presented. Section 4.3 shows chip implementation details and discusses the measurement results. Implementation of an ECG SoC based on the proposed microcontroller and the proposed memory block in the previous chapter is demonstrated in Section 4.4 as an application example. Conclusion is drawn in Section 4.5.

4.2 Subthreshold Application Specific 16-Bit RISC Microcontroller Architecture

A new 16-bit application-specific microcontroller which includes DSP-like instruction sets such as multiply-accumulate, and DSP-like addressing modes such as auto-increment and auto-decrement is presented in this section. The proposed architecture allows efficient implementation of biosignal processing algorithms including filter, data compression, and more. The architecture is kept concise and efficient in order to reduce power for power-limited sensor applications. The proposed core has a RISC instruction set which includes 33 instructions and 6 different addressing modes as summarizes in Table 4.1. All instructions are 16-bit wide and executed every one clock cycle i.e. average clock-per-instruction (CPI) is 1. The microcontroller has three general purpose and one indirect addressing registers. All internal registers and data paths are 16 bits and ALU operations are performed on 16-bit data.

Table 4.1. Instruction set and addressing modes of the proposed biomedical microcontroller core.

Instruction set		Addressing Modes
Arithmetic	ADD, ADDC, SUB, SUBC, CMP, MUL, MAC	Register (R1, R2, R3) Immediate Direct Indirect Indirect-Auto-Increment Indirect-Auto-Decrement
Logical	AND, OR, XOR, INV, RRA, RRC, RLA, RLC, TEST	
Load and Store	MOV, LOADL, LOADH, LOAD, IN, OUT, PUSH, POP	
Branch	CALL, RET, RETI, JMP, JZ, JC, JP, JV	
Special	SLEEP	

The core features Harvard architecture and a carefully-designed 4-stage pipeline as shown in the microarchitecture diagram in Fig. 4.1. This microarchitecture is designed in such a way that it prevents any hazard or stall in the pipeline and executes all branches, calls, and interrupts in one cycle. The number of stages in the pipeline depends on the type of instruction. All arithmetic operations will go through the instruction fetch (IF), decode (DE) and data memory access, execute (EX), and write-back (WB) stages. Every stage is completed within a half clock cycle. The custom-designed asynchronous SRAM allows data access during a half period of the clock. If the result of one arithmetic operation is going to be used in the next instruction, a fast-forward path bypasses the write-back stage and transfers the result of the previous instruction directly to the decode stage of the next instruction. All other instructions need only the first two stages to be executed as can be seen from the lower section of Fig. 4.1. This flexible configuration allows saving power during the operation modes like transferring blocks of data, by turning off the whole ALU via a power switch.

The proposed instruction set architecture (ISA) has been evaluated by

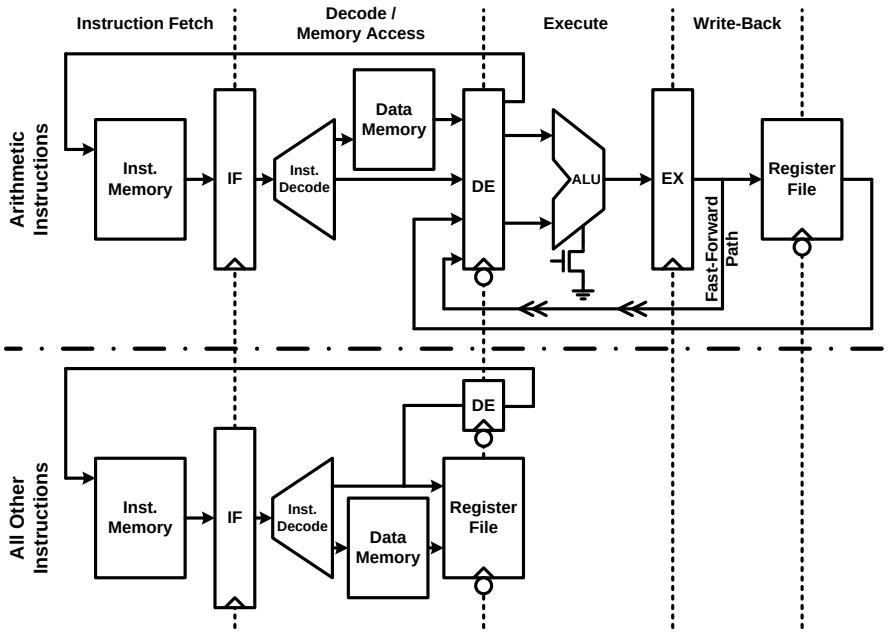


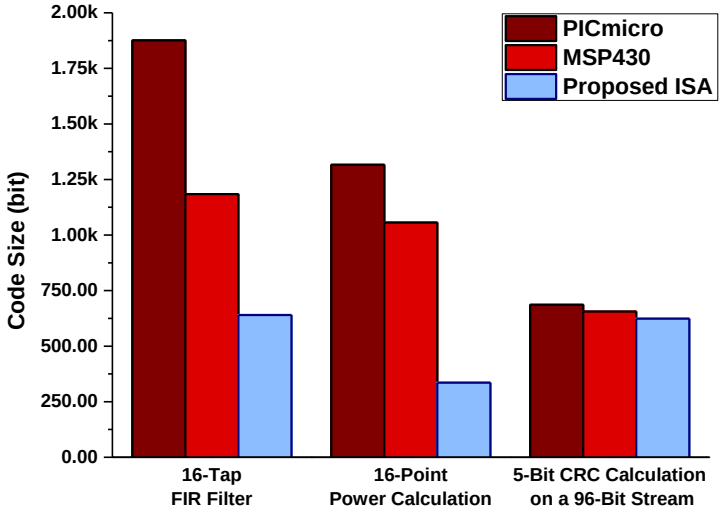
Fig. 4.1. Microarchitecture of the proposed microcontroller core.

comparing its performance against widely-used ultra-low-power architectures like PICmicro from Microchip [40] and MSP430 from Texas Instruments [45]. PICmicro has an 8-bit RISC architecture with 49 instructions and 3 addressing modes while MSP430 features a 16-bit RISC architecture with 27 instructions and 7 addressing modes. For the purpose of this comparison, various biomedical applications were reviewed and three commonly-used tasks in the digital back-end of such applications were selected. The first selected task was filtering. Filters are used to smooth the signal, select or reject a specific frequency band [65], or as a matched filter to detect a pattern [66]. In this comparison, a 16-tap finite impulse response (FIR) filter was implemented in each microcontroller. Power calculation was selected as the second task. Calculating average power of the signal over a window is required in various feature extraction algorithms [65]. Therefore, an average power calculation routine over a 16-sample window was implemented in each microcontroller family. Finally, error detection is required during transmission of a signal from the sensor node

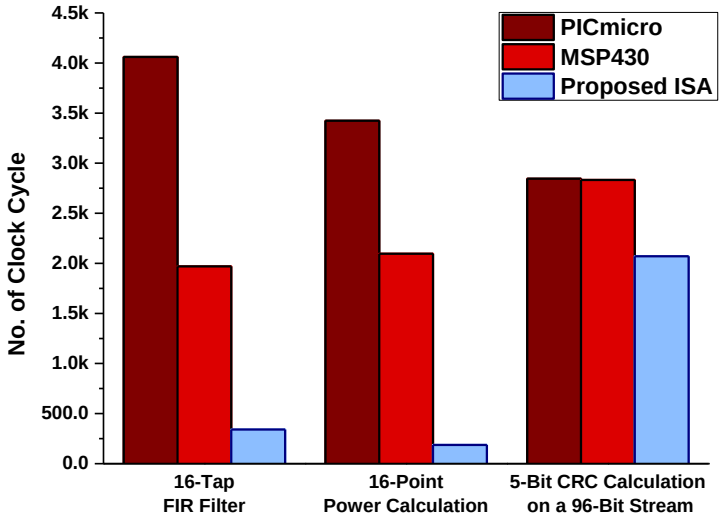
to the base station. Therefore the cyclic redundancy check (CRC), as an effective error detection algorithm, was selected as the third task. The 5-bit CRC which is used in the radio frequency identification (RFID) systems [67] was implemented as it is intended for very-low-power applications. These three routines are a fair representation of the whole digital back-end link from preprocessing and feature extraction to signal transmission.

All tasks were implemented in the PICmicro, MSP430 and the proposed ISA. 12-bit signed data was considered for all tasks as it is a common and acceptable bit size for biomedical applications. The implemented routines were examined for the code size and the number of clock cycles required to execute the task. As the instruction width is different among these microcontrollers, the code size in terms of the number of bits was used in the comparison. Smaller code size implies that a smaller code memory can be used in the system which translates to less leakage and active power in the memory. Fig. 4.2(a) shows the comparison of code size of the implemented tasks in three microcontrollers. As can be seen, the proposed instruction set considerably improves the code size in computational intensive tasks like filtering and power calculation.

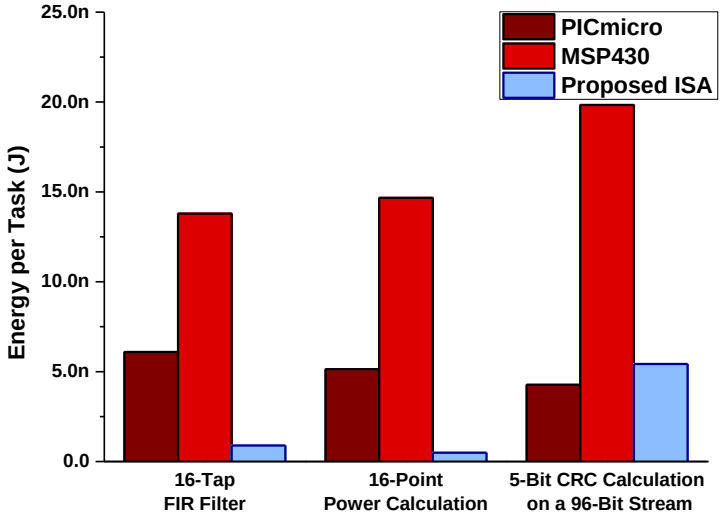
Next, the number of clock cycles required to complete one iteration of a task was evaluated for each routine. For example, in the FIR filter, filtering one sample, i.e. performing all 16 multiplications and additions for a 16-tap filter, was considered as one iteration. Fig. 4.2(b) shows the number of clock cycles for each task across three microcontrollers. According to this graph, the proposed architecture substantially reduces the number of clock cycles in each task. Given a fixed operating frequency for three microcontrollers, the fastest



(a)



(b)



(c)

Fig. 4.2. Comparison of (a) code size (b) number of clock cycles (c) energy per task of three different tasks, implemented using PICmicro, MSP430, and the proposed microcontroller architecture.

execution time is achieved by the proposed architecture. This reduction in the number of clock cycles and code size is due to adding the DSP features into the proposed architecture. To better examine the efficiency of the execution of the given tasks, the required energy for each task should be compared. This comparison will reveal any overhead due to complicated instructions or addressing modes. Since the proposed microcontroller operates in the subthreshold regime, for a fair comparison, two subthreshold implementations of PIC-micro [41] and MSP430 [46] were selected. The minimum energy per cycle was used to calculate energy per each task. Fig. 4.2(c) shows energy consumption during executing each task for three microcontrollers. As can be seen from this graph, while introducing complicated instructions in the proposed architecture helps greatly in reducing clock per task, the instruction set is sufficiently concise to allow very low energy per instruction. As a result, the proposed instruction set architecture allows implementing computationally-expensive tasks like filtering and power calculation with 6.8–28 times less energy compared to the other microcontroller families. For less computational tasks like CRC calculation, the performance of the proposed microcontroller is comparable to the performance of the minimal 8-bit microcontroller i.e. the added overhead is negligible.

The microcontroller is designed for sub/near-threshold operation to provide the best energy efficiency. To prevent any hold violations in the presence of high variations in the subthreshold regime, a dual-edge pipelining was implemented in which two consecutive stages are working with two opposite edges of the clock.

Fig. 4.3 shows the peripheral blocks implemented in the microcontroller.

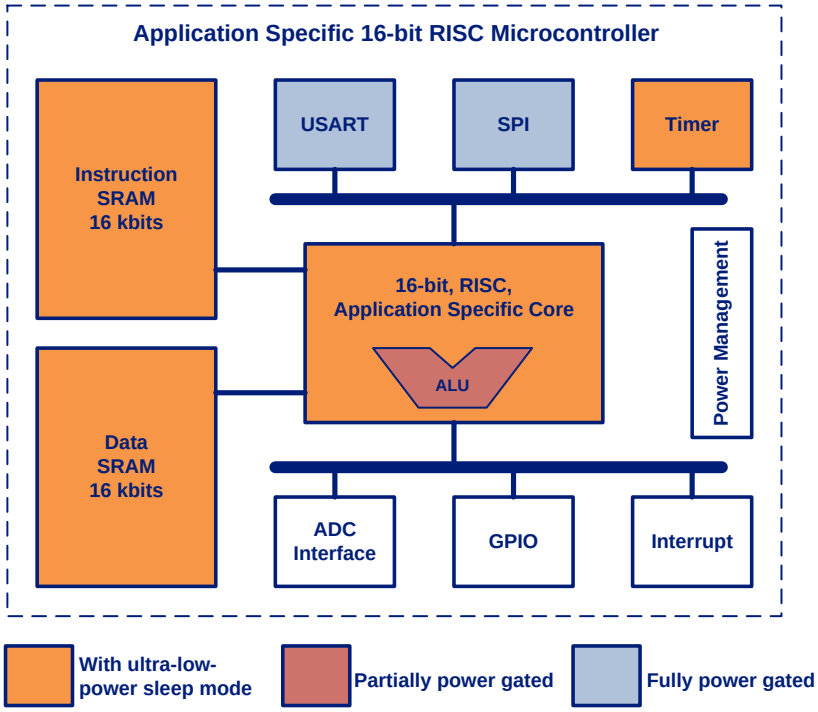


Fig. 4.3. Block diagram of the proposed application-specific biomedical microcontroller.

Two 16 kb custom-designed asynchronous SRAMs have been implemented to store code and data information. The power and speed of the code and data memory blocks are very important in the system as their power sets the minimum power level for the digital block and their speed limits the maximum operating frequency and the minimum energy-per-operation. These memory blocks were designed to work at the same voltage as the digital core, i.e. down to 0.25 V, to have the best power performance. To obtain high-speed performance and area efficiency at the subthreshold regime, the proposed architecture in Chapter 3 has been used to implement the memories of this microcontroller. Each 16 kb memory block has 32 columns and 512 rows

To have an efficient data transmission via RF link, a universal synchronous/asynchronous serial port (USART) is incorporated to generate serial data with an adjustable data rate which is useful in communication protocols. Other peripherals like timer, general purpose IO, and SPI are also included.

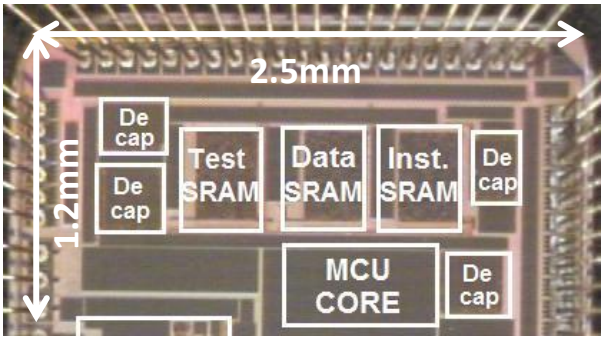


Fig. 4.4. Die photo of the fabricated microcontroller.

As the power consumption of sensor nodes is of main concern, extensive power and clock gatings are inserted in the microcontroller. First, larger peripherals like USART and SPI are power gated allowing complete shutdown while the system is in the standby mode or they are not being used. The ALU is also power gated for lowering power consumption while the microcontroller is idle or just transferring blocks of data. Second, a sleep mode, via a SLEEP command, is designed for the core and memory blocks. Under this mode, the whole core is clock gated and both memories are put into an ultra-low-power mode without losing their data.

4.3 Chip Implementation and Measurement Results

A prototype of the proposed microcontroller has been fabricated in a standard 0.13-μm bulk CMOS process. Fig. 4.4 shows the die photo of the chip. In this photo, the test SRAM, data SRAM, and inst. SRAM are three copies of the proposed SRAM in the previous chapter. The decaps are ordinary decoupling capacitor. The microcontroller and memories are fully functional down to 0.25 V supply voltage on average and 0.22 V in the best case as illustrated in the distribution of the minimum operating voltage in Fig. 4.5. In the current version of the system, a programmer writes the code into the code memory of the system via a parallel port at the first power-up. The average of

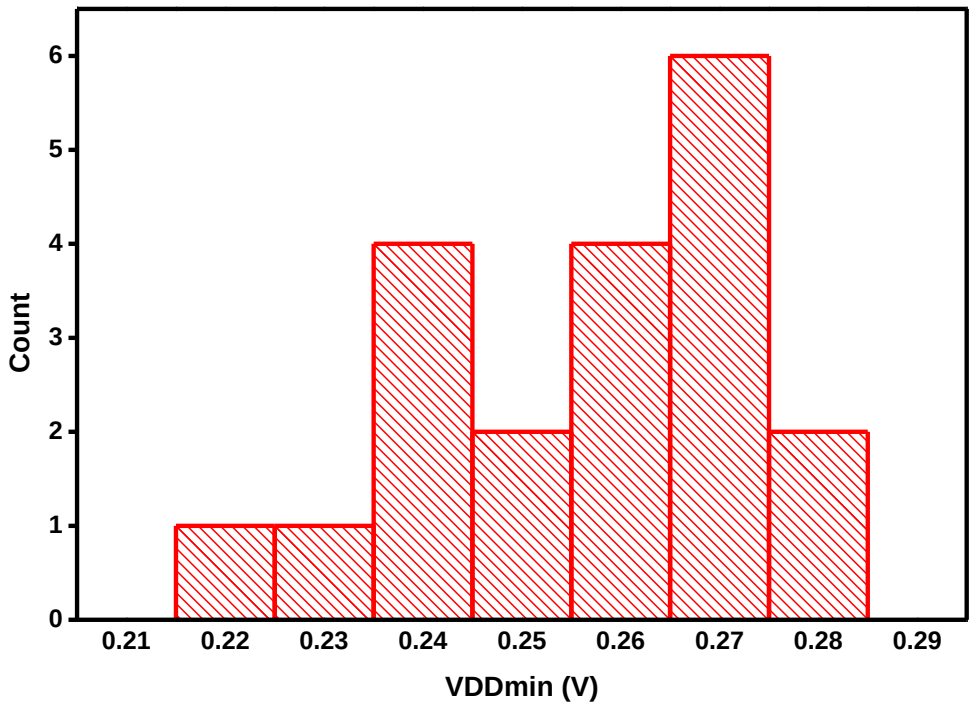


Fig. 4.5. Distribution of the minimum operating supply voltage for 20 chips.

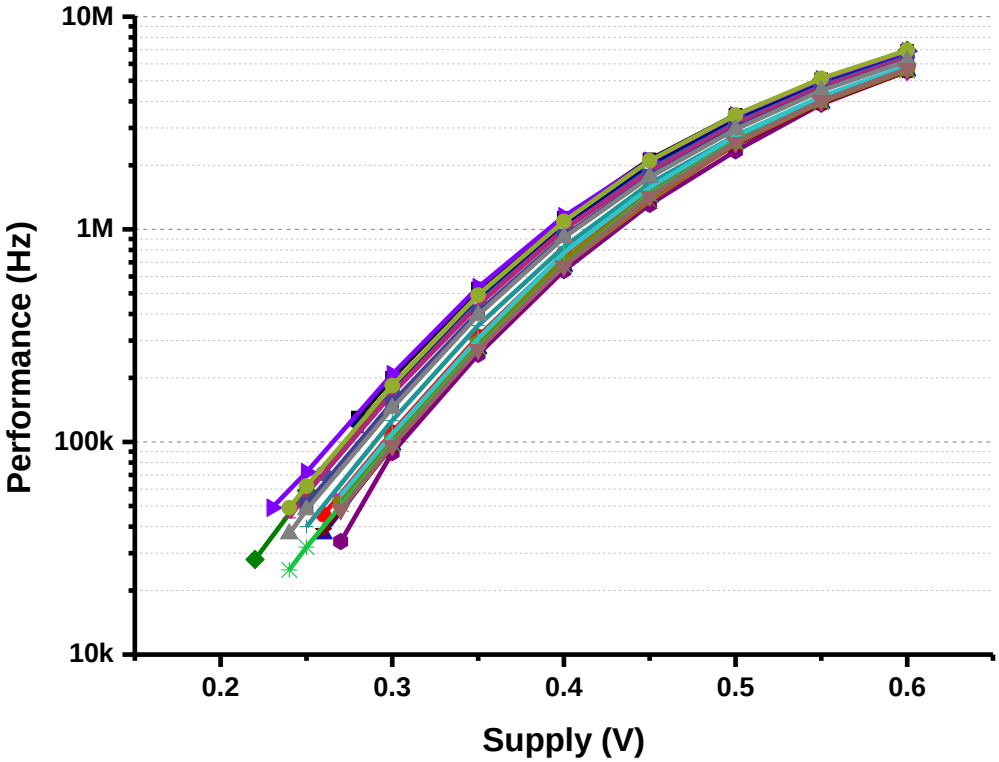


Fig. 4.6. Performance of the digital back-end across various supply voltages, measured for 20 chips.

the minimum supply voltage is 0.256 V with 17 mV standard deviation. All chips are fully functional at 0.28 V. Fig. 4.6 shows the maximum operating frequency of the microcontroller at various supply voltages for 20 chips. As

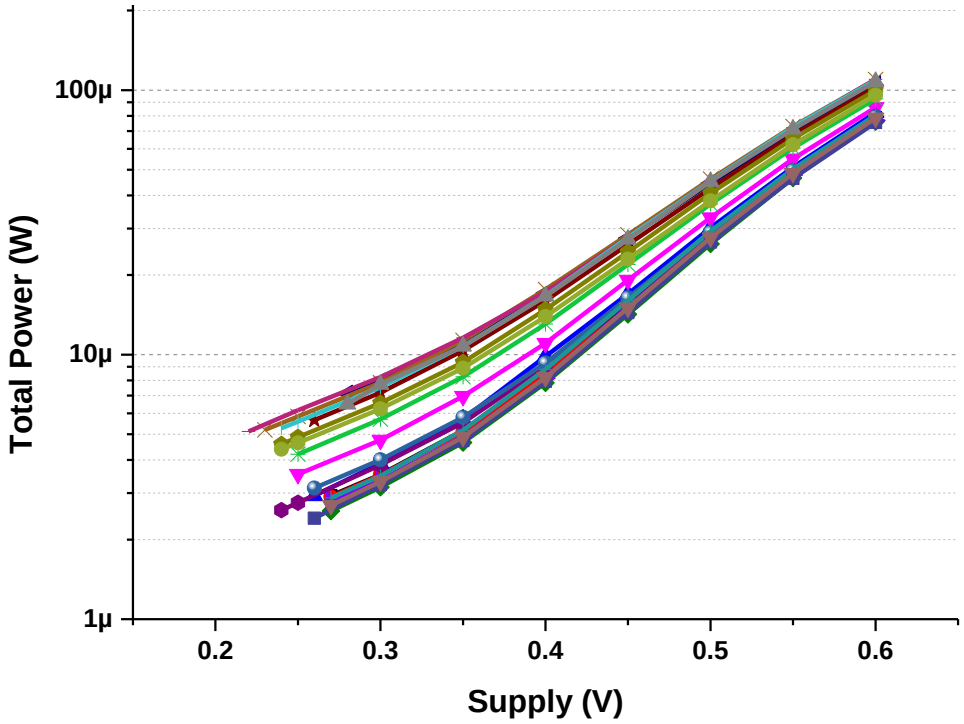


Fig. 4.7. Total power consumption of the digital back-end at full speed across various supply voltages, measured for 20 chips.

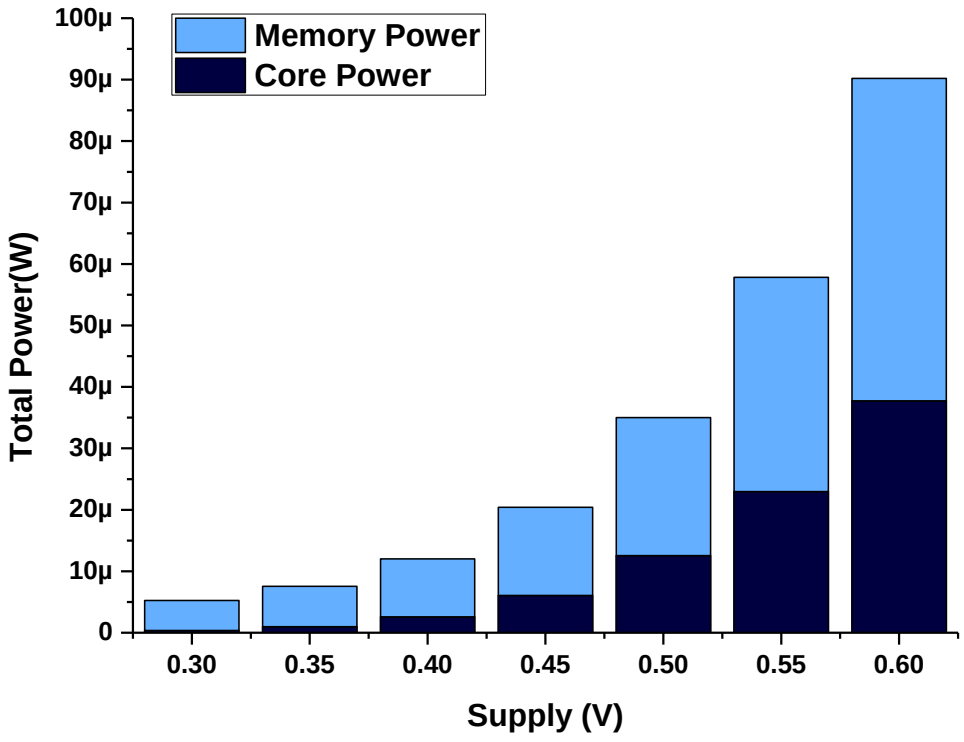


Fig. 4.8. The breakdown of the total power consumption in terms of memory and core power consumption across various supply voltages.

can be seen, the microcontroller is working from 52 kHz at 0.25 V to 6.1 MHz at 0.6 V. At the full speed, it consumes from 4.5 µW at 0.25 V to 90.2 µW at 0.6 V as shown in Fig. 4.7. These power consumption values include power of

the instruction and data SRAM. The contributions of the SRAM and micro-controller core to the total power are depicted in Fig. 4.8. At higher supply voltages, like 0.6 V, around 60% of the total power is consumed by the SRAM blocks while at voltages below threshold, like 0.3 V, almost all power is consumed by leakage power of the SRAM while the core power is negligible. This phenomenon would be worse in more advanced technologies in which their leakage current is much higher. This graph clearly shows the importance of designing ultra-low-power low-leakage SRAM in power-sensitive applications.

Leakage power of the digital back-end and its variation is depicted in Fig. 4.9. As expected, high variations in the sub/near-threshold regime causes around 3x variation in leakage power. In terms of leakage breakdown, most of leakage power is consumed by the SRAM blocks as shown in Fig. 4.10.

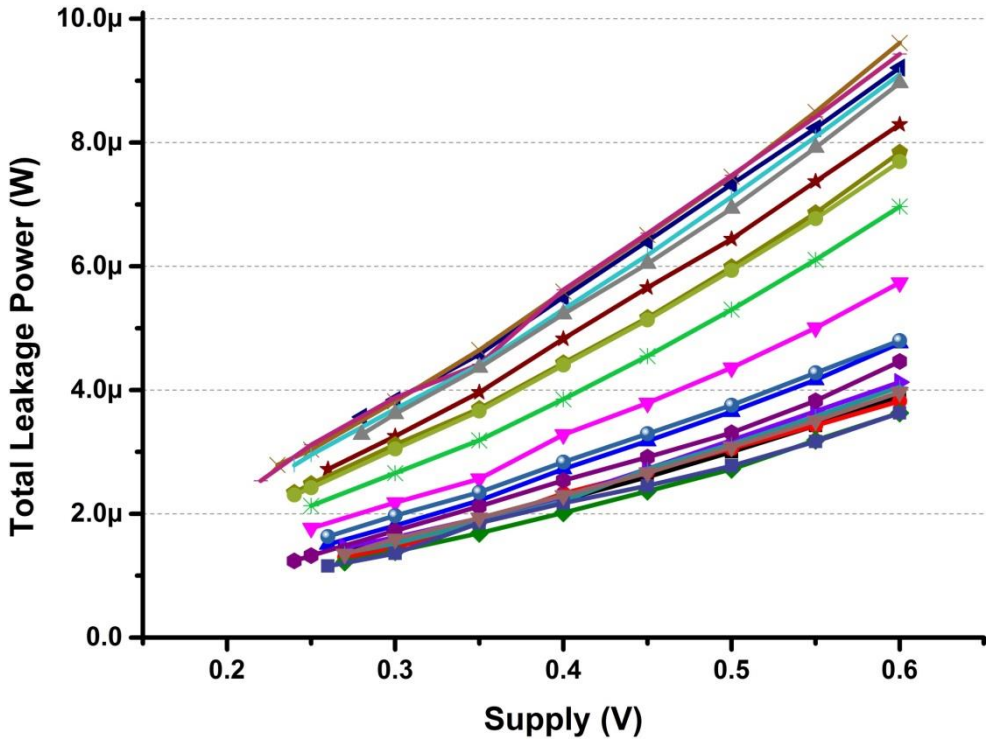


Fig. 4.9. Total leakage power of the digital back-end across various supply voltages, measured for 20 chips.

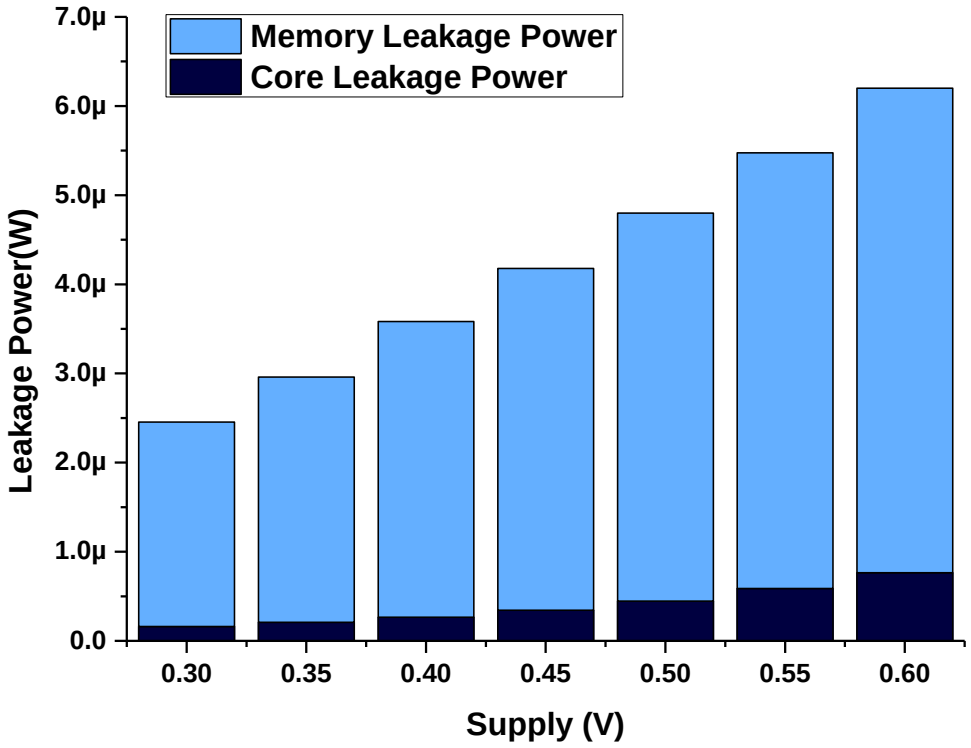


Fig. 4.10. The breakdown of the total leakage power in terms of memory and core leakage power across various supply voltages.

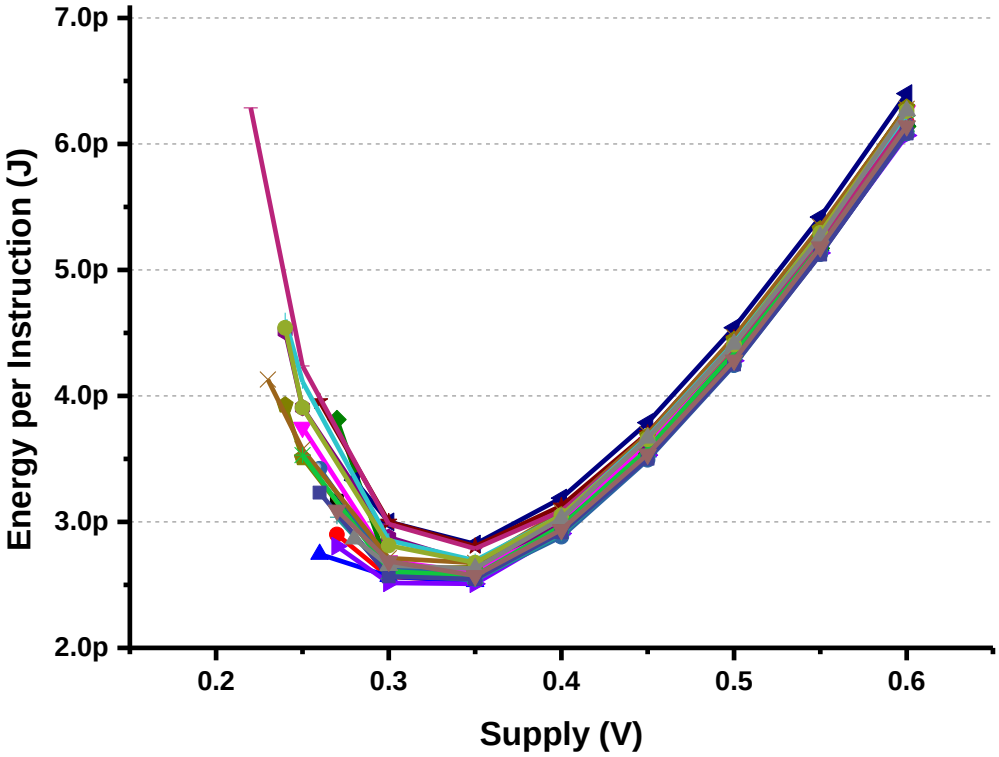


Fig. 4.11. Average energy-per-instruction of the microcontroller core across various supply voltages, measured for 20 chips.

Fig. 4.11 shows the average energy-per-instruction for the microcontroller core. At the minimum-energy point, the microcontroller consumes only

2.62 pJ per instruction on average. As predicted, this point is located near-threshold, i.e. 0.35 V. Considering the 16-bit core and complicated instructions like multiply and multiply-accumulate, this core shows interesting energy consumption per instruction. Table 4.2 summarizes the specifications and performance of the digital back-end of the fabricated SoC and compares it with the recent state-of-the-art works. It should be noted that the values in this table are the average performance of 20 chips. This work achieves the lowest operating voltage and the fastest performance while gives promising average energy-per-instruction.

Table 4.2. Specifications and performance summary and comparison.

	VLSIsymp11 [62]	ISSCC12 [63]	JSSC09 [46]	This work
Technology	0.18 μm	0.13 μm	65 nm	0.13 μm
Min Supply	1.2 V	0.3 V	0.3 V	0.25 V
Memory size	5 kB + 4 kB + 32 kB	5.5 kB	128 kb	16 kb + 16 kb
Microcontroller Architecture	32-bit, SIMD, Custom-designed	8-bit, RISC, PIC micro	16-bit, RISC, MSP430	16-bit, RISC, Custom-designed
Energy/Inst.	N.A.	1.5 pJ @ 0.5 V	~7 pJ @ 0.5 V	2.62 pJ @ 0.35 V
Operating Frequency	1 MHz @ 1.2 V	2 kHz @ 0.3 V 1.7 MHz @ 0.6 V	8.7 kHz @ 0.3 V 1 MHz @ 0.6 V	52 kHz @ 0.25 V 6.1 MHz @ 0.6 V
Total Power (Core + SRAM)	~12 μW	2.1 μW	~11.8 μW @ 0.5 V	4.5 μW @ 0.25 V 90.2 μW @ 0.6 V
Leakage power (Core + SRAM)	N.A.	N.A.	1 μW @ 0.3 V	2.3 μW @ 0.25 V 6.2 μW @ 0.6 V

4.4 Application Example: 3-Lead Wireless ECG System-on-Chip

The functionality and performance of the proposed microcontroller has been further evaluated by implementing it in a single-chip wireless electrocardiogram (ECG). The block diagram of the wireless ECG platform is shown in Fig. 4.12. The system front-end has two fully-differential channels for a typical 3-lead ECG recording, which performs signal filtering and amplification

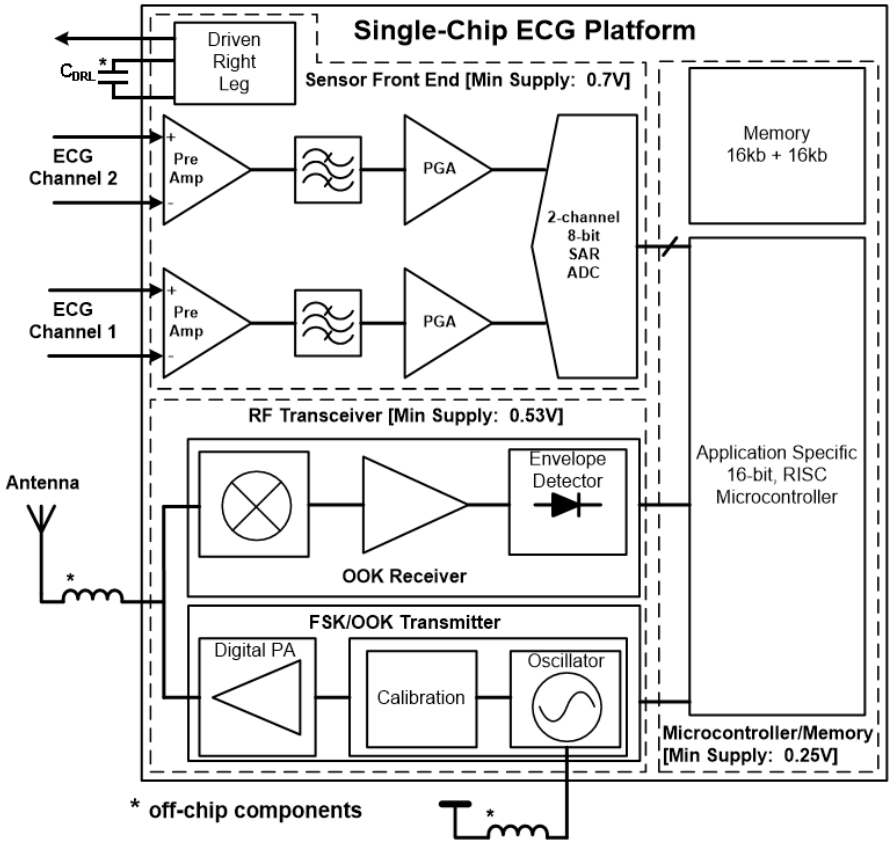


Fig. 4.12. Block diagram of the implemented single-chip ECG platform.

for ECG inputs before the quantization. A driven-right-leg (DRL) circuit is implemented to reduce common-mode interferences, especially from 50-Hz or 60-Hz power-line noises. A 2-channel 8-bit SAR ADC provides simultaneous sampling for both channels but quantizes them sequentially. Multiplexing after quantization removes any cross-talk between two channels. Simultaneous sampling is necessary to construct the third lead ECG signal based on the signals from the other two leads. The output of the ADC is connected to the proposed 16-bit RISC microcontroller. To facilitate an effective communication between the ADC and microcontroller, an interrupt port is dedicated to indicate the end of the AD conversion. Two implemented banks of 16 kb ultra-low-power subthreshold SRAMs facilitate feature extraction and lossless data compression before transmitting via an RF link by buffering ECG signals, leading to significant reduction in the amount of transmitted raw data. In

addition, duty cycling of the RF link is implemented to further optimize power consumption of the entire system. The proposed SRAM is working at the same voltage as the rest of the digital core, e.g. 0.7-V, which eliminates the high supply voltage and level shifters at the interfaces.

The microcontroller is connected to a Medical Implantation Communication Service (MICS) band transmitter via a configurable serial port to send raw or processed ECG data to a gateway. The serial port allows both synchronous and asynchronous transmissions. The transmitter needs only two off-chip inductors, which minimize the cost of external components. Furthermore, the SoC also integrates a receiver to ensure a reliable communication between the sensor and a gateway via the RF link. As the amount of data in the transmit path is much higher than the receive path, a high-data rate FSK transmitter is used at the transmitter while a low-data rate OOK receiver is implemented in the receiving path.

To achieve ultra-low power, the proposed SoC is designed to operate at low voltage. For single-supply operation, the SoC can work under a 0.7-V supply. The digital sub-systems, i.e. the microcontroller and SRAMs, are fully functional with a supply as low as 0.25 V. The RF transceiver is functional at 0.53 V. It should be noted that the design of the analog front-end, ADC, and RF transceiver is beyond the scope of this dissertation and has been performed by other colleagues in our lab.

A prototype of the wireless ECG SoC has been fabricated in a standard 0.13-μm bulk CMOS process. Fig. 4.13 shows the die photo and floorplan of the chip. The total area of the chip is 2.4 mm × 2.5 mm, which is pad-limited.

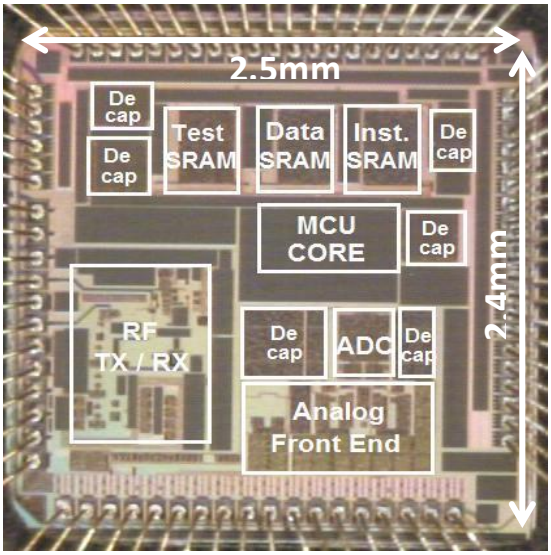


Fig. 4.13. Die photo and floorplan of the fabricated ECG platform.

Fig. 4.14 shows a real 3-lead ECG signal recording on a volunteer and compares it with a measurement from a commercial device. In this recording, Leads I and III were recorded while Lead II was derived accordingly. For the best energy efficiency, each block was working at its minimum supply voltage which can give the required performance. Therefore the digital back-end was working at 0.4 V and around 1 MHz, analog front-end and RF transceiver was working at 0.53 V and ADC was working at 0.7 V. The result shows good recording quality. During the recording, the whole system consumes 74.8 μW in the raw data transmission mode in which data is stored in the on-chip memory for 2 seconds (500 Hz sampling rate for both channels) and sent at the maximum bit-rate in a burst mode which gives less than 10% duty cycle for the RF transmission. In this mode, more than 80% of power is consumed by the RF transceiver and around 18% is consumed by the digital back-end. In the heart rate detection mode, the amount of transmitted data is much less than the raw data transmission mode. By transmitting heart rate information in every 5 seconds, this mode allows 0.5% duty cycling of the RF transmission which

leads to 17.4 µW power consumption of the system. In this mode 77% of power is consumed by the digital back-end and only around 17% is consumed by the RF transceiver. Table 4.3 summarizes specifications of the fabricated wireless ECG SoC and compares it with the recent state-of-the-art works.

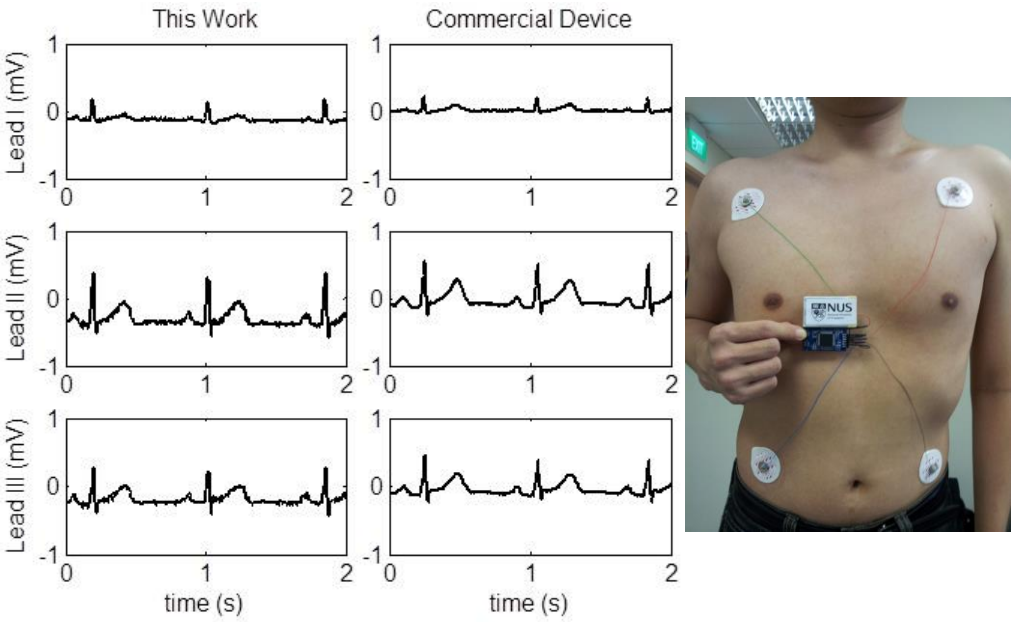


Fig. 4.14. Real 3-Lead ECG recording with no post-processing compared with a measurement from a commercial device.

Table 4.3. Wireless ECG SoC specifications and comparison.

	VLSIsymp11 [62]	ISSCC12 [63]	This work
Technology	0.18 µm	0.13 µm	0.13 µm
Area	5 mm × 4.7 mm	N.A.	2.4 mm × 2.5 mm
ECG Channels	3	4	2 (3 Leads)
Driven-Right-Leg	No	No	Yes
RF Transmitter	No	Yes	Yes
RF Receiver	No	No	Yes
Supply Voltages	1.2 V	0.3, 0.5, 1, 1.2 V (Harvesting)	0.7 V Single or 0.4, 0.53. 0.7 V Multiple
Total Power (Raw Data)	38 µW (No RF)	397 µW	74.8 µW
Total Power (Heart Rate)	31.1 µW (No RF)	19 µW	17.4 µW

4.5 Summary

In this chapter a novel 16-bit RISC microcontroller has been proposed for biomedical applications. The added DSP-like instructions and addressing modes make it an ideal choice for low-power computational-intensive tasks like biomedical algorithms. The silicon measurement proves functionality down to 0.25 V and a minimum energy point of 2.62 pJ per instruction at 0.35 V. The proposed architecture gives the lowest energy consumption and the fastest performance for common tasks like filtering. A 3-lead wireless ECG SoC was implemented based on the proposed microcontroller architecture. The implemented SoC proved the effectiveness of the proposed architecture in reducing total power consumption of the whole system.

Chapter 5
Conclusion and Future Works

5.1 Conclusion

Power consumption arguably is the most important design factor in the recent millimeter-scale systems like sensor nodes, wearable and implantable biomedical devices in which energy efficiency is of main concern. This dissertation focused on reducing power consumption of the digital subsection of a system-on-chip. Operation in the sub/near-threshold regime was chosen for the best energy consumption; however design challenges in this regime needed to be addressed.

An introduction on the whole topic and the scope of this work was first provided followed by a comprehensive review of the recent state-of-the-art works. This review was started with an overview on SRAM basics and assist techniques for the subthreshold operation. Then, various SRAM cells which

have been proposed by other works in the recent literatures for operation in the sub/near-threshold regime were reviewed. The challenges in this area were also highlighted. Lastly, the recent works in designing subthreshold microcontrollers were also reviewed.

Second, the challenges in designing a SRAM in the sub/near-threshold regime were addressed by a new SRAM architecture, named average-8T write/read-decoupled cell. This architecture provided data-independent leakage, differential read, and pseudo-write tolerance features. Furthermore, it allowed optimization between the minimum operating voltage and area. Performance of this architecture was verified by fabricating and testing two 64 kb and 16 kb memory blocks in the 0.13 µm bulk CMOS process. In addition, usage of the reverse narrow-width effect (RNWE) in sizing of the SRAM transistors in the subthreshold regime was examined. Another 16 kb memory block in the 65 nm bulk CMOS technology was fabricated and tested based on RNWE-aware sizing and the average-8T architecture. The measurement results verified effectiveness of this sizing technique.

Third, towards a full system, a microcontroller core with few carefully-selected DSP features was designed. The designed architecture substantially reduced memory size, the number of clock cycle, and total energy required for computational intensive tasks like filtering. A subthreshold implementation of this architecture was designed and fabricated in the 0.13 µm bulk CMOS process. The memory designed in the first part of this work was used as instruction and data memories of the implemented microcontroller. The whole microcontroller was functional at a single supply down to 0.25 V.

Lastly, as an application example, the designed microcontroller and memory were integrated with an analog front-end, a SAR ADC, and an RF transceiver to make a single-chip 3-lead wireless ECG SoC. The fabricated chip was fully functional down to 0.7 V due to the limitation of ADC. The digital back-end could work down to 0.25 V. The fabricated SoC achieved the lowest power consumption in the heart-rate detection and raw data transmission modes among other reported biomedical SoCs.

5.2 Future Works

In the next step, this work can be continued in various directions. First of all, further reduction in leakage power of the SRAM is required. According to the results of Section 4.4, at low voltages leakage power of the SRAM is the dominant portion of the total power. Therefore, SRAM leakage reduction leads to the total system power reduction. Second, variation in the subthreshold regime can be addressed by adding tuneability to the logic. In the SRAM, a tunable sense amplifier and read or write timing can be helpful. In digital logics like the microcontroller, error-resiliency can be implemented to mitigate the effect of variations.

Third, the microcontroller architecture for biomedical applications can be further investigated. Instruction sets based on other architectures like very long instruction word (VLIW) and more complicated instructions and addressing modes can be studied.

Lastly, an energy scavenging block and a power management unit (PMU) should be added to the proposed SoC to have a complete energy-autonomous system. A very-low-power and precise clock generator circuit is

also required for such a system.

Bibliography

[1] P. Fiorini, I. Doms, C. Van Hoof, and R. Vullers, "Micropower energy scavenging," in *IEEE European Solid-State Circuits Conf. (ESSCIRC) Dig. Tech. Papers*, pp. 4-9, Sep. 2008.

[2] A. Wang, and A. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310-319, Jan. 2005.

[3] G. Chen, M. Fojtik, D. Kim, D. Fick, J. Park, M. Seok, M.-T. Chen, Z. Foo, D. Sylvester, and D. Blaauw, "Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 288-289, Feb. 2010.

[4] A. P. Chandrakasan, D. C. Daly, J. Kwong, and Y. K. Ramadass, "Next generation micro-power systems," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 2-5, Jun. 2008.

[5] S. R. Sridhara, M. DiRenzo, S. Lingam, S.-J. Lee, R. Blazquez, J. Maxey, S. Ghanem, Y.-H. Lee, R. Abdallah, P. Singh, and M. Goel, "Microwatt embedded processor platform for medical system-on-chip applications," *IEEE J. Solid-State Circuits*, vol. 46, no. 4, pp. 721-730, Apr. 2011.

[6] A. Wang, B. H. Calhoun, and A. P. Chandrakasan, *Subthreshold Design for Ultra Low-Power Systems*, New York: Springer, 2006.

[7] J. Chen, L. T. Clark, and T.-H. Chen, "An ultra-low-power memory with

a subthreshold power supply voltage," *IEEE J. Solid-State Circuits,* vol. 41, no. 10, pp. 2344-2353, Oct. 2006.

[8] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "A variation-tolerant sub-200 mV 6-T subthreshold SRAM," *IEEE J. Solid-State Circuits,* vol. 43, no. 10, pp. 2338-2348, Oct. 2008.

[9] M.-E. Hwang, and K. Roy, "A 135mV 0.13µW process tolerant 6T subthreshold DTMOS SRAM in 90nm technology," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, pp. 419-422, Sept. 2008.

[10] L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, L. Sekaric, S. J. McNab, A. W. Topol, C. D. Adams, K. W. Guarini, and W. Haensch, "Stable SRAM cell design for the 32 nm node and beyond," in *Symp. VLSI Technology Dig. Tech. Papers*, pp. 128-129, Jun. 2005.

[11] N. Verma, and A. P. Chandrakasan, "A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy," *IEEE J. Solid-State Circuits,* vol. 43, no. 1, pp. 141-149, Jan. 2008.

[12] M. E. Sinangil, N. Verma, and A. P. Chandrakasan, "A reconfigurable 8T ultra-dynamic voltage scalable (U-DVS) SRAM in 65 nm CMOS," *IEEE J. Solid-State Circuits,* vol. 44, no. 11, pp. 3163-3173, Nov. 2009.

[13] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, Y. Nakase, and H. Shinohara, "A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 158-159, Jun. 2009.

[14] Z. Liu, and V. Kursun, "Characterization of a novel nine-transistor SRAM cell," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.,* vol. 16, no. 4, pp. 488-492, Apr. 2008.

[15] S. Lutkemeier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Porrmann, and U. Ruckert, "A 65 nm 32 b subthreshold processor with 9T multi-Vt SRAM and adaptive supply voltage control," *IEEE J. Solid-State Circuits,* vol. 48, no. 1, pp. 8-19, Jan. 2013.

[16] J.-J. Wu, Y.-H. Chen, M.-F. Chang, P.-W. Chou, C.-Y. Chen, H.-J. Liao, M.-B. Chen, Y.-H. Chu, W.-C. Wu, and H. Yamauchi, "A large σVTH/VDD tolerant zigzag 8T SRAM with area-efficient decoupled differential sensing and fast write-back scheme," *IEEE J. Solid-State Circuits,* vol. 46, no. 4, pp. 815-827, Apr. 2011.

[17] J. P. Kulkarni, A. Goel, P. Ndai, and K. Roy, "A read-disturb-free, differential sensing 1R/1W port, 8T bitcell array," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.,* vol. 19, no. 9, pp. 1727-1730, Sep. 2011.

[18] B. H. Calhoun, and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *IEEE J. Solid-State Circuits,* vol. 42, no. 3, pp. 680-688, Mar. 2007.

[19] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A 0.2 V, 480 kb subthreshold SRAM with 1 k cells per bitline for ultra-low-voltage computing," *IEEE J. Solid-State Circuits,* vol. 43, no. 2, pp. 518-529, Feb. 2008.

[20] I. J. Chang, J.-J. Kim, S. P. Park, and K. Roy, "A 32 kb 10T sub-threshold SRAM array with bit-Interleaving and differential read scheme in 90 nm CMOS," *IEEE J. Solid-State Circuits,* vol. 44, no. 2, pp. 650-658, Feb. 2009.

[21] S. Clerc, F. Abouzeid, G. Gasiot, D. Gauthier, and P. Roche, "A 65nm SRAM achieving 250mV retention and 350mV, 1MHz, 55fJ/bit access energy, with bit-interleaved radiation Soft Error tolerance," in *IEEE European Solid-State Circuits Conf. (ESSCIRC) Dig. Tech. Papers*, pp. 313-316, Sep. 2012.

[22] M.-F. Chang, J.-J. Wu, K.-T. Chen, Y.-C. Chen, Y.-H. Chen, R. Lee, H.-J. Liao, and H. Yamauchi, "A differential data-aware power-supplied (D^2AP) 8T SRAM cell with expanded write/read stabilities for lower VDDmin applications," *IEEE J. Solid-State Circuits,* vol. 45, no. 6, pp. 1234-1245, Jun. 2010.

[23] J. P. Kulkarni, K. Keejong, and K. Roy, "A 160 mV robust Schmitt trigger based subthreshold SRAM," *IEEE J. Solid-State Circuits,* vol. 42, no. 10, pp. 2303-2313, Oct. 2007.

[24] J. P. Kulkarni, and K. Roy, "Ultralow-voltage process-variation-tolerant schmitt-trigger-based SRAM design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.,* vol. 20, no. 2, pp. 319 - 332, Feb. 2012.

[25] M.-F. Chang, S.-W. Chang, P.-W. Chou, and W.-C. Wu, "A 130 mV SRAM with expanded write and read margins for subthreshold applications," *IEEE J. Solid-State Circuits,* vol. 46, no. 2, pp. 520-529,

Feb. 2011.

[26] R. V. Joshi, R. Kanj, and V. Ramadurai, "A novel column-decoupled 8T cell for low-power differential and domino-based SRAM design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 5, pp. 869-882, May 2011.

[27] D. Anh-Tuan, J. Y. S. Low, J. Y. L. Low, Z.-H. Kong, X. Tan, and K.-S. Yeo, "An 8T differential SRAM with improved noise margin for bit-interleaving in 65 nm CMOS," *IEEE Trans. Circuits Syst. I, Reg. Papers,* vol. 58, no. 6, pp. 1252-1263, Jun. 2011.

[28] M.-H. Tu, J.-Y. Lin, M.-C. Tsai, S.-J. Jou, and C.-T. Chuang, "Single-ended subthreshold SRAM with asymmetrical write/read-assist," *IEEE Trans. Circuits Syst. I, Reg. Papers,* vol. 57, no. 12, pp. 3039-3047, Dec. 2010.

[29] C. Ming-Pin, C. Lai-Fu, C. Meng-Fan, Y. Shu-Meng, K. Yao-Jen, W. Jui-Jen, H. Mon-Shu, S. Hsiu-Yun, C. Yuan-Hua, W. Wen-Ching, Y. Tzu-Yi, and H. Yamauchi, "A 260mV L-shaped 7T SRAM with bit-line (BL) swing expansion schemes based on boosted BL, asymmetric-Vth read-port, and offset cell VDD biasing techniques," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 112-113, June 2012.

[30] M.-H. Tu, J.-Y. Lin, M.-C. Tsai, C.-Y. Lu, Y.-J. Lin, M.-H. Wang, H.-S. Huang, K.-D. Lee, W.-C. Shih, S.-J. Jou, and C.-T. Chuang, "A Single-Ended Disturb-Free 9T Subthreshold SRAM With Cross-Point Data-Aware Write Word-Line Structure, Negative Bit-Line, and Adaptive

Read Operation Timing Tracing," *IEEE J. Solid-State Circuits,* vol. 47, no. 6, pp. 1469-1482, Jun. 2012.

[31] L. Chang, R. K. Montoye, Y. Nakamura, K. A. Batson, R. J. Eickemeyer, R. H. Dennard, W. Haensch, and D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE J. Solid-State Circuits,* vol. 43, no. 4, pp. 956-963, Apr. 2008.

[32] T.-H. Kim, J. Liu, and C. H. Kim, "A voltage scalable 0.26 V, 64 kb 8T SRAM with V_{min} lowering techniques and deep sleep mode," *IEEE J. Solid-State Circuits,* vol. 44, no. 6, pp. 1785-1795, Jun. 2009.

[33] B. H. Calhoun, and A. Chandrakasan, "A 256kb sub-threshold SRAM in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 628-629, Feb. 2006.

[34] L. Nazhandali, B. Zhai, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, T. Austin, and D. Blaauw, "Energy optimization of subthreshold-voltage sensor network processors," in *Proc. Int. Symp. Computer Architecture (ISCA),* pp. 197-207, 4-8 June 2005.

[35] B. Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson, A. Reeves, M. Minuth, R. Helfand, T. Austin, D. Sylvester, and D. Blaauw, "Energy-efficient subthreshold processor design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 8, pp. 1127-1137, Aug. 2009.

[36] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, D. Blaauw, and T. Austin, "A 2.60pJ/inst subthreshold sensor

processor for optimal energy efficiency," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 154-155, Jun. 2006.

[37] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, "Performance and variability optimization strategies in a sub-200mV, 3.5pJ/inst, 11nW subthreshold processor," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 152-153, Jun. 2007.

[38] B. Zhai, R. G. Dreslinski, D. Blaauw, T. Mudge, and D. Sylvester, "Energy efficient near-threshold chip multi-processing," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, pp. 32-37, Aug. 2007.

[39] M. Seok, S. Hanson, Y.-S. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "The Phoenix Processor: A 30pW platform for sensor applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 188-189, Jun. 2008.

[40] PICmicro. Microchip. [Online]. Avialable: http://ww1.microchip.com/downloads/en/devicedoc/33023a.pdf.

[41] S. C. Jocke, J. F. Bolus, S. N. Wooters, A. D. Jurik, A. C. Weaver, T. N. Blalock, and B. H. Calhoun, "A 2.6uW sub-threshold mixed-signal ECG SoC," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 60-61, Jun. 2009.

[42] H. Kaul, M. A. Anders, S. K. Mathew, S. K. Hsu, A. Agarwal, R. K. Krishnamurthy, and S. Borkar, "A 320 mV 56 uW 411 GOPS/Watt ultra-low voltage motion estimation accelerator in 65 nm CMOS," *IEEE J.*

Solid-State Circuits, vol. 44, no. 1, pp. 107-114, Jan. 2009.

[43] Y. Pu, J. P. de Gyvez, H. Corporaal, and Y. Ha, "An ultra-low-energy/frame multi-standard JPEG co-processor in 65nm CMOS with sub/near-threshold power supply," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 146-147,147a, Feb. 2009.

[44] M. Seok, D. Jeon, C. Chakrabarti, D. Blaauw, and D. Sylvester, "A 0.27V 30MHz 17.7nJ/transform 1024-pt complex FFT core with super-pipelining," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 342-344, Feb. 2011.

[45] MSP430. Texas Instruments. [Online]. Avialable: http://www.ti.com/lit/ug/slau056k/slau056k.pdf.

[46] J. Kwong, Y. K. Ramadass, N. Verma, and A. P. Chandrakasan, "A 65 nm sub-vt microcontroller with integrated SRAM and switched capacitor DC-DC converter," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 115-126, Jan. 2009.

[47] J. Kwong, and A. P. Chandrakasan, "An energy-efficient biomedical signal processing platform," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1742-1753, Jul. 2011.

[48] D. Bol, J. De Vos, C. Hocquet, F. Botman, F. Durvaux, S. Boyd, D. Flandre, and J. Legat, "A 25MHz 7uW/MHz ultra-low-voltage microcontroller SoC in 65nm LP/GP CMOS for low-carbon wireless sensor nodes," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 490-492, Feb. 2012.

[49] ARM Cortex-M3. ARM. [Online]. Avialable: http://www.arm.com/products/processors/cortex-m/cortex-m3.php.

[50] A. Agarwal, S. K. Mathew, S. K. Hsu, M. A. Anders, H. Kaul, F. Sheikh, R. Ramanarayanan, S. Srinivasan, R. Krishnamurthy, and S. Borkar, "A 320mV-to-1.2V on-die fine-grained reconfigurable fabric for DSP/media accelerators in 32nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 328-329, Feb. 2010.

[51] M. Ashouei, J. Hulzink, M. Konijnenburg, Z. Jun, F. Duarte, A. Breeschoten, J. Huisken, J. Stuyt, H. De Groot, F. Barat, J. David, and J. Van Ginderdeuren, "A voltage-scalable biomedical signal processor running ECG using 13pJ/cycle at 1MHz and 0.4V," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 332-334, Feb. 2011.

[52] CoolFlux. NXP. [Online]. Avialable: http://www.coolflux.com/.

[53] N. Ickes, G. Gammie, M. E. Sinangil, R. Rithe, J. Gu, A. Wang, H. Mair, S. R. Datla, B. Rong, S. Honnavara-Prasad, L. Ho, G. Baldwin, D. Buss, A. P. Chandrakasan, and U. Ko, "A 28 nm 0.6 V low power DSP for mobile applications," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 35-46, Jan. 2012.

[54] N. Ickes, Y. Sinangil, F. Pappalardo, E. Guidetti, and A. P. Chandrakasan, "A 10 pJ/cycle ultra-low-voltage 32-bit microprocessor system-on-chip," in *IEEE European Solid-State Circuits Conf. (ESSCIRC) Dig. Tech. Papers*, pp. 159-162, Sep. 2011.

[55] S. Luetkemeier, T. Jungeblut, M. Porrmann, and U. Rueckert, "A 200mV

32b subthreshold processor with adaptive supply voltage control," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 484-486, Feb. 2012.

[56] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits,* vol. 22, no. 5, pp. 748-754, Oct. 1987.

[57] J. Wang, S. Nalam, and B. H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, pp. 129-134, Aug. 2008.

[58] T.-H. Kim, J. Liu, and C. H. Kim, "An 8T Subthreshold SRAM Cell Utilizing Reverse Short Channel Effect for Write Margin and Read Performance Improvement," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, pp. 241-244, Sep. 2007.

[59] N. Shigyo, and T. Hiraoka, "A review of narrow-channel effects for STI MOSFET's: A difference between surface- and buried-channel cases," *Solid-State Electronics,* vol. 43, no. 11, pp. 2061-2066, Nov. 1999.

[60] M.-H. Chang, Y.-T. Chiu, S.-L. Lai, and W. Hwang, "A 1kb 9T subthreshold SRAM with bit-interleaving scheme in 65nm CMOS," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, pp. 291-296, Aug. 2011.

[61] X. Zou, X. Xu, L. Yao, and Y. Lian, "A 1-V 450-nW fully integrated programmable biomedical sensor interface chip," *IEEE J. Solid-State Circuits,* vol. 44, no. 4, pp. 1067-1077, Apr. 2009.

[62] H. Kim, R. F. Yazicioglu, S. Kim, N. V. Helleputte, A. Artes, M. Konijnenburg, J. Huisken, J. Penders, and C. V. Hoof, "A configurable and low-power mixed signal SoC for portable ECG monitoring applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, pp. 142-143, Jun. 2011.

[63] F. Zhang, Y. Zhang, J. Silver, Y. Shakhsheer, M. Nagaraju, A. Klinefelter, J. Pandey, J. Boley, E. Carlson, A. Shrivastava, B. Otis, and B. Calhoun, "A batteryless 19uW MICS/ISM-band energy harvesting body area sensor node SoC," in *Solid-State Circuits Conference (ISSCC), IEEE International*, pp. 298-300, Feb. 2012.

[64] J. Hulzink, M. Konijnenburg, M. Ashouei, A. Breeschoten, T. Berset, J. Huisken, J. Stuyt, H. de Groot, F. Barat, J. David, and J. Van Ginderdeuren, "An ultra low energy biomedical signal processing system operating at near-threshold," *IEEE Trans. Biomedical Circuits Syst. (TBCAS)*, vol. 5, no. 6, pp. 546-554, Dec. 2011.

[65] K. Kim, U. Cho, Y. Jung, and J. Kim, "Design and implementation of biomedical SoC for implantable cardioverter defibrillators," in *IEEE Asian Solid-State Circuits Conf. (A-SSCC) Dig. Tech. Papers*, pp. 248-251, Nov. 2007.

[66] A. Ruha, S. Sallinen, and S. Nissila, "A real-time microprocessor QRS detector system with a 1-ms timing accuracy for the measurement of ambulatory HRV," *IEEE Trans. Biomedical Engineering*, vol. 44, no. 3, pp. 159-167, Mar. 1997.

[67] *Information technology -- Radio frequency identification for item management -- Part 6: Parameters for air interface communications at 860 MHz to 960 MHz* International Organization for Standardization (ISO), 2010.